Essays on the Distributive Politics of Bureaucracy

Tara Slough

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

ABSTRACT

Essays on the Distributive Politics of Bureaucracy

Tara Slough

Bureaucrats are hired to produce public goods. Yet, despite the distributive implications of this canonical rationale, bureaucrats are generally absent from theories of "who gets what." The three papers in this dissertation advance a role for bureaucrats in the distribution of public goods and services premised on their work in policy implementation. I provide new theory and evidence to answer three questions. The first paper asks the question: under what conditions do bureaucrats' actions generate inequalities in the provision of public services? The second paper inquires: how does the design of bureaucratic oversight institutions influence a state's capacity to implement policy and citizen access to services? The third essay asks: how does the co-production of public goods by politicians and bureaucrats influence voters' ability to hold politicians to account?

In the first paper, I study the conditions under which bureaucratic bias (discrimination) in the allocation of services generates inequality in access. I argue that citizens' principal mechanism of control over bureaucrats is to complain to a politician. When politicians respond to complaints by tightening oversight of bureaucrats, differences in citizens' access to complain induce bureaucrats to devote more effort to groups with the loudest voices. I test this theory using a national-scale factorial audit experiment of Colombia's two largest national social welfare programs to measure bureaucratic effort behaviorally. I find that bureaucrats provide less information about social welfare programs to poor citizens and internal migrants. Consistent with the theory, such bias manifests most strongly in places with greater inequalities in citizens' ability to access the state and on tasks where oversight from politicians is most likely. These results are unlikely to reflect taste-based discrimination or screening. This paper shows that inequality in access to public goods and services can emerge even when politicians' budget allocations to public goods are equitable.

In the second paper, I examine the distributional consequences of the use of citizen complaints in bureaucratic oversight. I study the adoption and consequences of bureaucratic oversight institutions

in the context of service provision. Specifically, I consider a politician's choice to use (or ignore) information generated by complaints when monitoring a bureaucrat. Complaints generate information that direct a politicians' remediation of bureaucratic decisions and may increase bureaucratic effort. However, when costs of complaint vary across the population, the use of this information generates inequality in the distribution of service outputs, improving the access of citizens that can complain while reducing the access of citizens that cannot. Further, relying on citizen information can build or erode a state's capacity to accurately implement public policies, depending on the distribution of these costs across the population. This paper introduces citizen complaint systems as an institution that shapes both policy implementation capacity and distributional outcomes in comparative perspective.

In the final paper, I start from the observation that in many theories of electoral accountability, voters learn about an incumbent's quality through the observation of public goods outcomes. However, politicians rely on bureaucracies to produce public goods. Across contexts, politicians work with bureaucracies of markedly different qualities. In this paper, I argue that accountability relations between voters and politicians yield different empirical implications at different levels of bureaucratic quality. I introduce a model of electoral accountability with a voter, a politician, and a bureaucrat. The model identifies observational equivalencies between (i.) the implications of pooling equilibria that emerge at high and low levels of bureaucratic quality (with informed, rational voters) and (ii.) the findings of existing studies that are interpreted to indicate a lack of accountability due to uninformed or irrational voters. I demonstrate the plausibility of the model by introducing and validating an original measure of bureaucratic quality in Brazilian municipalities. I use this measure to extend four studies on corruption and accountability. I conclude with implications for the comparative study of accountability across the world's democracies.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation reflects the generosity and insight of many. These acknowledgments are not comprehensive; I only hope that those that I have neglected to mention by name know that I remember and am grateful for the their support.

I spent most of the decade before arriving in New York playing violin and started this degree with a vaguer idea than most about what I had signed up for. I remain thankful to the Columbia Political Science department for taking a chance on me and helping me to build a foundation from which to explore.

My greatest fortune as a grad student has been finding an incredible set of mentors and advisors who have supported me and my learning. First and foremost, I am grateful to my dissertation chair, Mike Ting. Mike has been exceedingly generous with his time for discussion and his careful feedback on my work. Conversations with Mike about the material in this dissertation and elsewhere improved my work and were a highlight of my time at Columbia. Given the clear differences in our research interests and approaches, working with Mike was anything but preordained, but I could not have asked for a better advisor.

In my first semester Comparative Politics survey, Macartan Humphreys shaped my ideals and expectations about what political science research can (and to my taste, should) look like. Macartan's ability to ask challenging questions that generate new ideas is unrivaled. Several pieces of the dissertation stem from such questions. In addition to his contributions to my thinking, Macartan has provided many opportunities for me outside the department at Columbia by connecting me with opportunities and partners. He also hosted me during visits to the WZB Social Science Center in Berlin.

John Huber encouraged the questions and approaches in this dissertation while always imploring me to remember the big picture. In my case, this is not an easy job and I am grateful for his patience. John's insistence on clear writing and communication has been immensely helpful. Shortcomings in this regard are mine alone and I will work to continue to improve in ways reflective of his feedback and his (written) example.

John Marshall consistently provides careful and detailed feedback on my papers and his feedback has improved my work substantially. John's door is usually open, which I too often abused. I was fortunate to have had John as a sounding board as this project developed. John also gave particularly active and actionable advice when I was on the job market. I hope to model John's generosity and helpfulness as an advisor in this next stage.

Don Green inspires my enthusiasm for experiments and careful, clever empirical research more generally. Don has provided feedback on many grant proposals, research designs, and papers. I am grateful for his thorough comments and sound advice. I walk away from conversations with Don with a strengthened conviction that claims of a trade-off between "substantive" knowledge and technique are spurious. Don inspires me to do better on both fronts.

Kimuli Kasara has been a particularly generous mentor. She endeavors to keep me disciplined as a comparativist and communicator, difficult as that task may be. She has read many drafts of research proposals and papers and given insightful, honest feedback. I have learned much about this business, where it falls short, and how it might be improved from frequent communication with Kimuli.

Miriam Golden has also mentored me from afar for much of grad school. I met Miriam as a second-year student and worked with her for a year later in grad school. Miriam has been very generous in socializing me to a broader network and giving me perspective on the field. I look forward to seeing more of Miriam in New York once travel is "back on."

Others at Columbia deserve special notice. In particular, three classes and many conversations with Ben Goodrich were instrumental in helping me to build a foundation in statistics and research design. Ben's patience and generosity helped me learn to teach myself new skills. Yotam Margalit, who is missed at Columbia, has been a mentor despite the distance. His ability to frame and market work is second to none, and I am grateful for all that he has taught me. Carlo Prato has answered big and small questions alike and gave particularly generous feedback on my job talk. Vicky Murillo offered me opportunities for leadership, particularly within the greater New York Latin American politics community. Conversations with and comments from Suresh Naidu, Shigeo Hirano, and Jack Snyder have generated ideas and improved my work.

I am also exceedingly grateful to my cohort and colleagues at Columbia. I could not have asked for a better cohort than Abhit Bhandari, Matt Carr, Jasper Cooper, Alicia Cooperman, Alex de la Paz, Kolby Hanson, Jacob Kopas, Erik Lin-Greenberg, Jorge Mangonnet, Renanah Miles, Dane Thorley, Joonseok Yang, Erin York, and (honorarily) Lindsay Dolan. I also thank Anna Wilke, Gosha Syunyaev, Alex Coppock, and so many others for thoughtful feedback and suggestions over the years. I am so excited to see where we go from here.

I thank Charles Angelucci for his input and thoughtful comments as an external reader for my defense. I have learned from presenting portions of this dissertation to many generous conference, seminar, and policy audiences. Beyond those already mentioned here, I thank Deborah Beim, Natalia Garbiras-Díaz, Germán Gieczewski, Navin Kartik, Horacio Larreguy, Evan Lieberman, Gareth Nellis, Pablo Querubín, Pia Raffler, Tesalia Rizzo, and Jake Shapiro for their comments on parts of this dissertation.

The research in this dissertation has been supported by many individuals and organizations. I am indebted to Juan Sebastian Leiva, Sofia Jaramillo, Yuliana Silva Contreras, Jiawei Fu, and Carolina Torreblanca for outstanding research assistance. In Colombia, I am also grateful to the Garcia-Arciniegas family, the Cabrera-Venegas family, the late Juliana Giraldo Bahamon and her family, the IPA country office in Bogotá, and many others for their hospitality, friendship, and advice.

The dissertation was financed by an NSF Doctoral Dissertation Research Improvement Grant (#1729334) and grants from J-PAL, Evidence in Governance and Politics (EGAP), the Center for the Study of Development Strategies, and the Columbia Department of Political Science. Additionally, my graduate study was supported by an NSF Graduate Research Fellowship, DGE-11-44155. I have been fortunate to spend the last two years as predoctoral fellow at EGAP "at" the University of California, Berkeley. Thanks especially to Cyrus Samii and Jaclyn Leaver for their collaboration and support in this position. This past year (until the pandemic), I worked from my new academic home at New York University. I am honored to join such a smart, insightful group of scholars. I am grateful to Sandy Gordon, Alastair Smith, and David Stasavage for facilitating this transitional year at NYU.

My final thanks go to my family and friends. My parents, Greg and Linda, paved my way to higher education, have consistently supported my educational pursuits and, most importantly, have loved me unconditionally. I am so lucky to have a close relationship with my grandmother, Barbara, who has shown enthusiasm throughout my education. I don't know that anyone is as relieved as she is that I am finally done with school. My brother, Patrick, is exceedingly patient and generous with my inane programming questions and, more importantly, he and Sara are supportive from afar. Finally, Beni helped me to open my academic and professional horizons beyond music a decade ago. Beni, I am so lucky to share this adventure with you.

Chapter 1

Introduction

In the last 30 years, social policy across Latin America has expanded dramatically to encompass vulnerable populations that formerly received little non-discretionary support from the state (De la O, 2015; Garay, 2016). In particular, the extension of conditional cash transfers (CCTs), pensions, and health care to millions of poor citizens in the urban periphery and rural areas represents a stark transformation in the relationship between these citizens and the state. Given the magnitude of socioeconomic inequality in the region, possible beneficiaries from these expansions of social policy comprise up to 80% of national populations, respectively (Portes and Hoffman, 2003).

One consistent observation across the region holds that implementation of these programs has been uneven (e.g., Soares, Ribas, and Osório, 2010; Camacho and Conover, 2011; Parker and Todd, 2017; Niedzwiecki, 2018). While modest in absolute terms, the benefits afforded by CCT programs, pensions, and health care are non-trivial for the populations that they intend to serve. As such, uneven implementation can generate substantial welfare losses when vulnerable populations are unable to access these benefits. Our present understanding of implementation has generally focused on the behavior of politicians (Camacho and Conover, 2011; De la O, 2015; Niedzwiecki, 2018) or potential beneficiaries (Parker and Todd, 2017). Yet, given the scale of and targeting inherent to these social programs, most administration is done by bureaucrats. Less is known about bureaucrats' role in implementation (or implementation failures), even as administration remains a salient concern to program architects.[1]

Consideration of bureaucrats' role in social program implementation makes clear that bureaucrats' actions influence "who gets what" from the state. Conceptualization of the distributive role of bureaucrats in the provision of public goods and services raises questions about interactions between bureaucrats and other, more frequently studied actors. In particular, I examine a set of strategic interactions between bureaucrats, political principals (elected politicians), and citizens. This setting motivates a large set of unanswered questions. For example, under what conditions does oversight of bureaucrats exacerbate or mitigate disparities in service delivery? What motivates politicians to engage in different forms of bureaucratic oversight? What recourse do citizens have to contest bureaucratic decision making or performance? When do bureaucrats provide cover for malfeasant politicians?

This dissertation develops theory and evidence on strategic interactions between politicians, bureaucrats, and citizens to develop a role for bureaucrats in the study of distributive politics. While the first paper is situated in the context of Latin American social policy, the dissertation is not a case study of these policies. Instead, the dissertation aims to speak more generally to bureaucrats' role in policy implementation by generating hypotheses that are applicable in different geographic and policy domains. Perhaps the best-developed application in existing literature examines policing, studying how variation in police officer behavior yields uneven implementation of policies with substantial – even life and death – welfare consequences (for recent examples in a large literature, see Ba, 2018; Streeter, 2019; Rivera and Ba, 2019; Magaloni, Franco-Vivanco, and Mello, 2020; Knox, Lowe, and Mummolo, 2020). The papers in this dissertation collectively advance an agenda

[1] This observation comes from a series of interviews with employees of the Departamento Nacional de Planeación and Prosperidad Social in Bogotá, Colombia in 2017 and 2018.

on the distributive consequences of policy implementation by bureaucrats.

1.1 Bureaucracy and the Politics of "Who Gets What"

This dissertation consists of three essays on the distributive politics of bureaucracy. It is motivated by a general paucity of bureaucrats in accounts of "who gets what" from the state. This omission stands at odds with the canonical – and inherently distributive – rationale for bureaucrats, which holds that bureaucrats are "hired to produce public goods." This dissertation examines bureaucrats' role in the implementation and delivery of public goods and services, deriving implications for distribution, efficiency, and equity of state-provided goods and services.

The distributive politics tends to "star" politicians in its efforts to understand patterns of taxation and spending. As Golden and Min (2013) write, "distributive policies [are] those that involve taxes and transfers, and in particular the decisions about allocations of government goods and services to identifiable localities or groups" (p. 74). Focusing on transfers, most existing work equates allocation of transfers by politicians with the outputs received by citizens.[2] We typically measure distribution in terms of budget allocations or the outputs (i.e., public goods) accessed by individuals or districts. The former measure can miss substantial slippage between funds allocated and ultimate outputs when assessing "who gets what." The latter measure may mislead inferences about politicians' allocative considerations. As such, I focus on bureaucrats' role in policy implementation – the mapping of a budget allocation into a policy output. This dissertation provides theory and evidence that equating allocations and outputs can produce misleading assessments of "who gets what" and why. Further, it demonstrates that distinguishing allocations from outputs is not only consequential for understanding of bureaucratic behavior; it also generates new predictions about the behavior of

[2] Work on taxation more frequently acknowledges fiscal capacity as a constraint on politicians with distributive consequences. There is not consensus about whether fiscal capacity is a measure of bureaucratic capability/output or a different technology altogether. For example, Kasara and Suryanarayan (2015) equate fiscal capacity and bureaucratic capacity. The bureaucratic nature of fiscal capacity is less clear in works such as Tilly (1990) and Besley and Persson (2010).

both politicians and voters.

The essays provide answers to three central questions. First, *under what conditions do bureaucrats' actions generate inequalities in the provision of public services?* I begin from the observation that bureaucrats produce and distribute public goods and services, with wide scope to influence "who gets what." I contend that citizens' principal mechanism of control over bureaucrats is the complaint to a politician. When politicians respond to complaints by tightening oversight of bureaucrats, differences in citizens' access to complain induce bureaucrats to devote more effort to groups with the loudest voices. I test this theory using a national-scale factorial audit experiment of Colombia's two largest national social welfare programs to measure bureaucratic effort behaviorally. I find that bureaucrats provide less information about social welfare programs to poor citizens and internal migrants. Consistent with the theory, I show that bias manifests most strongly in places with greater inequalities in citizens' ability to access the state and on tasks where oversight from politicians is most likely.

Second, *how do institutions stipulating bureaucratic oversight influence a state's capacity to implement policy and inequality in the distribution of public goods and services?* I start from a similar observation to the previous paper that bureaucratic oversight frequently relies upon information provided by citizen complaints. I propose a model of service provision to study the adoption and consequences of oversight institutions. Specifically, I consider a politician's choice to use (or ignore) information generated by complaints when monitoring a bureaucrat. Complaints generate information that direct a politicians' remediation of bureaucratic decisions and may increase bureaucratic effort. However, when costs of complaint vary across the population, use of this information generates inequality in the distribution of service outputs, improving access of citizens that can complain while reducing access of citizens that cannot. Further, relying on citizen information can build or erode a state's capacity for policy implementation, depending on the distribution of these costs across the population. In this paper, I introduce citizen complaint systems as an institution that shapes both policy implementation capacity and distributional outcomes in comparative perspective.

Third, *how does the co-production of public goods and services by politicians and bureaucrats influence voters' ability to hold politicians accountable? How does this co-production affect our ability to diagnose accountability failures empirically?* In many theories of electoral accountability, voters learn about an incumbent's quality through the observation of public goods outcomes. Globally, politicians rely on bureaucracies to produce public goods, but politicians across contexts work with bureaucracies of markedly different qualities. In this paper, I argue that accountability relations between voters and politicians yield different empirical implications at different levels bureaucratic quality. I introduce a model of electoral accountability with a voter, a politician and a bureaucrat. The model identifies observational equivalencies between the implications of pooling equilibria that emerge at high and low levels of bureaucratic quality (with informed, rational voters) and the findings of existing studies that are interpreted to indicate a lack of accountability due to uninformed or irrational voters. I demonstrate the plausibility of the model by introducing and validating an original measure of bureaucratic quality in Brazilian municipalities and using it to extend four studies on corruption and accountability. I conclude with implications for the comparative study of accountability across the world's democracies.

1.2 Organizing Insights

The essays in this dissertation are united by three overarching insights that represent a departure from modal approaches to the study of bureaucracy and comparative politics more generally. In the following discussion, I contextualize and justify these developments as they relate to the three essays.

1.2.1 Insight #1: Bureaucrats and Politicians Do Different Things

In each essay in this dissertation, politicians and bureaucrats "do different things." More specifically, the strategies available to bureaucrats and politicians are distinct. In the first essay, the bureaucrats distribute a service and a politician monitors the bureaucrat's decision. In the second essay, the politician designs a contract by which to monitor a bureaucrat's service provision to a citizen. In the final essay, the politician allocates a budget while a bureaucrat produces the public good with that budget.

Standard approaches to modeling the agency relationships between bureaucrats and politicians in the last two decades have emphasized the study of administrative policymaking (Gailmard and Patty, 2012). While the precise strategies available to bureaucrats and politicians do generally differ slightly (i.e., a politician could set policy or delegate the policy determination to the bureaucrat), ultimately, politicians and bureaucrats do very similar things with respect to policymaking. Furthermore, both are motivated by by ideological preferences over policy. Studying ideological conflict between politicians and bureaucrats in this setting has generated many insights. For example, new ideas about delegation have emerged from this body of literature (e.g., Epstein and O'Halloran, 1994; Gailmard and Patty, 2007; Lewis, 2008; Fox and Jordan, 2011). Yet, these models ultimately focus on one task of bureaucrats among many. In so doing, models of this sort implicitly restrict attention to a comparatively small set of high-level bureaucrats who participate in policymaking. Empirically, the vast majority of bureaucrats (ideally) do not have the power to make policy. Moreover, the focus on *ideological* preferences of politicians and bureaucrats further limits scope to polities where ideology is more salient as an organizing feature of politics. The latter concern simply reflects the challenges inherent to transportation of a US-centric literature to other settings.[3]

[3]Comparative applications of delegation models of this class including Huber and Shipan (2002); McCarty (2004); Huber and McCarty (2004) are important contributions to this literature, but represent a small proportion of models of this type.

The approach I adopt in this dissertation provides a complementary perspective. By studying domains in which politicians and bureaucrats do different things, I move from a focus on bureaucratic policymaking to bureaucratic policy implementation. This move ultimately broadens attention to the actions of a larger set of bureaucrats. For example, in the Colombian study (Chapter 2), I measure the behavior of the public servants that enroll citizens in two social programs in municipalities across the country. This stands in contrast to the heads of the agencies running these programs who have the power to determine some aspects of social policy design. Obviously, these local program administrators individually have less ability to influence outcomes than the respective agency directors. However, I contend that in the aggregate, their actions have substantial implications for access. As such, this work reorients our focus to a set of bureaucrats that is emergent in the empirical literature (Rasul and Rogger, 2016; Brierly, 2020; Raffler, 2019), but underrepresented in existing theories about bureaucracy in Political Science.

1.2.2 Insight #2: Bureaucrats, Politicians, and Citizens

The models advanced in these papers characterize strategic relationships between a politician, a bureaucrat, and a citizen.[4] The papers incorporate all three actors by nesting two frequently-studied strategic interactions. First, they consider accountability relationships between citizens and politicians. The papers collectively study two forms of accountability relations. Chapter 4 studies accountability relations in an electoral setting (Fearon, 1999; Ashworth, 2012). Chapters 2 and 3 study a more quotidian manifestation of these relationships, that of citizen complaints. The study of complaint- or claim-making as a form of widespread engagement between voters and politicians has grown in recent years, particularly in the context of developing democracies (Kruks-Wisner, 2018; Bussell, 2019). These interactions mark an regular feature of constituency service by politicians.

Second, I examine moral hazard problems in bureaucracies. The canonical framework empha-

[4] In the electoral context in Chapter 4, the "citizen" is a "voter."

sizes the interactions between politicians and bureaucrats (or principals and agents) in which the politician cannot (fully) observe the actions of the bureaucrat. While the literature is replete with accounts of moral hazard, the papers in this dissertation join a small group of papers that integrate both voter-politician accountability relations with studies of moral hazard in bureaucracies. Chapter 2 and Chapter 3 draw substantial inspiration from the emphasis on citizen complaint in bureaucracies put forth by Prendergast (2003, 2007). Chapter 4 joins a small set of models of electoral accountability with a bureaucrat (Fox and Jordan, 2011; Yazaki, 2018; Li, Sasso, and Turner, 2019). I draw observable implications from these theories and present new data measuring (across chapters) the behavior of politicians, bureaucrats, and citizens to expand this nascent body of literature.

1.2.3 Insight #3: On the Study of "Bad Politics"

In conducting preliminary fieldwork for this dissertation in Colombia, I talked to a number of public servants. When I introduced myself, I started by explaining that I was a PhD student studying "bureaucracy." An early response was a stated aversion to the bureaucrat/bureaucracy language. The word "bureaucrat," I was told, conjured images of a lazy, inept, or corrupt government employee. The polite language was "public servant" (*servidor público*) and "public administration" (*función pública*), respectively. I changed my pitch accordingly.

The study of bureaucrats in low- and middle-income countries would suggest that my interlocutors' fear of being characterized as lazy, corrupt, or inept is well-founded. Dominant themes in the study of developing country bureaucracies emphasize patronage-laden cronyism and petty corruption (Leff, 1964; Mauro, 1995; Van Rijckeghem and Weder, 2001; Rasul and Rogger, 2016; Brierly, 2020). While there is ample evidence that these features present in some bureaucracies, we have much less evidence on the prevalence of cronyism and corruption across countries, agencies, or even bureaucrats. Moreover, we do not know whether these features are the primary drivers of inefficiency or welfare loss.

This characterization of bureaucracies is reflective of what I refer to as "bad politics" frame for the study of politics. In this frame, scholars look to an actor with "bad" motives that map directly onto inefficient or normatively undesirable outcomes. The literature is replete with examples of venal politicians, corrupt bureaucrats, and know-nothing voters. It is, after all, easy to identify examples of individuals that personify these descriptions. Yet, an immediate embrace of "bad" motives as drivers of bad outcomes absent other possible explanations limits our understanding of why bad outcomes emerge.

"Bad politics" weakens the project of Comparative Politics, in particular, because of the asymmetry in the frequency of its application across regions. Bad outcomes (as we define them) are more prevalent in the Global South. When we overemphasize bad motives explanations of these outcomes, we cede unnecessary ground in explaining similarities and differences of political phenomena across polities. For example, consider the comparatively benign treatment of pork in American Politics vis-a-vis work studying the scourge of targeted redistribution (i.e., clientelism) in developing democracies.[5] Similarly, while debates about voters' capacity for rational retrospection continue globally, we implicitly or explicitly attribute worse outcomes to uninformed or irrational voters in regions where politician malfeasance is (traditionally) more apparent. For example, when American Politics scholars provide evidence of limits to rational voter retrospection in the United States (Campbell et al., 1960; Healy, Malhotra, and Mo, 2010; Achen and Bartels, 2016), the consequences – in terms of political outcomes – are never as dire as the consequences attributed to low-information voting in the Global South (Dunning et al., 2019).

The observation of "bad politics" alongside its asymmetric application is far from new. In a

[5]The *quid pro quo* exchanges signified by the term "clientelism" could be argued to be normatively worse than other forms of targeted distribution. Yet, evidence of *quid pro quo* exchanges is exceedingly rare in the empirical literature (Hicken and Nathan, 2020). The distinction between forms of targeted distribution across contexts seems, to me, to be weaker than generally conveyed.

blistering critique of a 1969 book drawing comparisons between Colombian and US politicians, Hirschman (1970: p. 331) summarized the author's argument, writing:

> "Colombian politicians are selfish..., ambitious, unscrupulous, unprincipled, exceedingly demagogic-interested exclusively in increasing their own power, always ready to betray yesterday's friends and allies, and, to top it all, incapable of having friendly personal relations with anyone because they feel comfortable only with abject supplicants... On the other hand, there is the politician with a [genuine interest in programs and policies] whose preferred habitat is the United States of America. He enjoys working on concrete policies and achieving a stated goal; hence he is principled, willing to defend unpopular causes, always ready to come to constructive agreements, hardworking, and generally lovable."

Certainly, there is now more subtlety in the description of politics of developing countries and greater skepticism about the benevolence of United States politicians than when Hirschman wrote fifty years ago. Hirschman and others have proposed a variety of remedies to such issues in the study of Comparative Politics. The remedies I attempt to embrace in this dissertation focus on developing theories that are both symmetrically applicable across contexts and which advance mechanisms for bad outcomes that are distinct from "bad politics."

First, I intend for the theories in each chapter to be symmetrically applicable across contexts. While the evidence is largely drawn from single cases, namely Colombia and Brazil, the theories offer predictions as to why outcomes may differ across contexts. Chapters 2 and 3 emphasize differences in the underlying population as a determinant of levels of service provision. Chapter 3 further considers how oversight institutions may yield different outcomes across contexts as a function of existing public personnel systems. Chapter 4 derives predictions for distinct manifestations of electoral accountability at different levels of bureaucratic quality.

Second, I endeavor to step away from “bad politics” explanations for bad outcomes. The dissertation studies three phenomena that would likely be classified as normatively bad political outcomes: bias in access to state services; capture of the state and erosion of its ability to serve the population; and corruption and underinvestment in public goods. In contrast to many explanations for these phenomena, I do not assert the presence of “bad politics” in the form of an ill-intentioned political actor seeking to produce a bad outcome. Specifically, I do not rely on necessarily venal politicians, corrupt bureaucrats, or uninformed voters to generate these outcomes.[6] This is not an argument that any actor maximizes welfare: they do not. Instead, I study a series of agency problems that can generate these sub-optimal outcomes.

My argument is not that “bad politics” explanations are wrong or not useful. I simply posit that these accounts must be evaluated against other explanations that generate similar patterns of outcomes. The design of policies or interventions to improve outcomes depends on the mechanism that generates such outcomes. As such, there are very practical justifications for considering a wider class of potential mechanisms.

1.3 Approach and Methods

1.3.1 Theory and Empirics

The papers in this dissertation reflect the conviction that theory and empirical work can productively be integrated. Each paper contains a game theoretic model and the empirical papers draw inspiration from the identification (or credibility) revolution. The idea that these two approaches can productively be combined has been a matter of debate over the last decade (Clark and Golder, 2015; Ashworth, Berry, and de Mesquita, 2015; Samii, 2016; Huber, 2017). The move toward

[6]The models in Chapters 2 and 4 do allow for some aspects of “bad politics.” Chapter 2 allows for bureaucrats or politicians with biased tastes and Chapter 4 allows for uninformed voters. In the empirical applications, however, I do not find evidence in favor of these manifestations of “bad politics” in the contexts that I study.

design-based inference has focused scholars' energy on the assumptions invoked to identify select causal estimands. These assumptions are typically viewed as empirical, not theoretical. In other work, I argue that in a wide subset of applications, basic theoretical assumptions govern whether oft-estimated causal estimands are defined (Slough, 2019). In this sense, the models help to support claims of causal identification. In keeping with the design-based motivation, with one exception, all of the empirical results reported in this dissertation are estimates of reduced-form quantities (with respect to the models).[7]

Beyond considerations of identification, situating a research design within an equilibrium offers much greater clarity about what is being estimated empirically (under the assumptions of a model). In design-based research, mapping the treatment assignment onto the model provides implications for whether a given causal quantity tests an implication of interest (Bueno de Mesquita and Tyson, 2019). It also guides assessment of which empirical comparisons (observational or experimental) "make sense" in light of a researcher's model of the world and which comparisons do not "make sense." Finally, the models give clarity on the question of whether causal estimands estimate equilibrium or partial-equilibrium effects. The distinction is not necessarily obvious in the absence of a model and helps to clarify *why* we observe the effects, which is particularly important for informing policy.

1.3.2 Original vs. Secondary Data

Finally, the dissertation as a whole reflects multiple approaches toward the data collection and analysis. As a field, we often place a premium on the collection of original data. Importantly, these incentives push researchers to tackle new questions and and pursue novel answers. However, it becomes much more challenging to adjudicate between theories/arguments when each theory is tested in its own context with its own data. One approach to this problem, captured by the push for im-

[7]The structural parameter Δ in Chapter 2 also admits a clear reduced-form interpretation.

proved external validity, emphasizes replication of studies in multiple contexts, as in the Metaketa initiative by Evidence in Governance and Politics (EGAP). In either case, the tendency to resort to "getting new data" as a discipline imposes costs, reinforces some inequities, and, in the case of human subjects research, raises some ethical questions. In working on these papers, I have taken the view that a slight reduction of emphasis on original data in favor of secondary analysis of existing data may be productive. The secondary analyses here are made possible by the generosity and transparency of the authors that collected the original data. The broader disciplinary move toward reproducibility is certainly helpful in this regard.

In Chapter 2, I document the results of an original audit experiment conducted in Colombia. The motivation was to directly measure bureaucratic effort in a matter not captured by existing measures. I complement the original experimental data with several sources of administrative data collected from various Colombian government agencies to probe the mechanisms underlying bureaucratic bias. I designed the experiment in consultation with government agencies overseeing the two social programs I audit and worked to minimize the costs and harms of the experiment to subjects through design decisions enumerated in more detail in Chapter 2. In sum, I endeavor to justify for the value of the experiment and data collection where measures were not previously available.

In Chapter 4, I take a very different approach to data collection. I introduce and test new theory using secondary analyses of several existing studies. In particular, I develop a new measure of bureaucratic quality from administrative data to adjudicate between my theory and existing arguments in order to demonstrate the plausibility of the model. The literature information and accountability in low- and middle-income democracies has grown quite large in the last 15 years. The growing availability of reproducible data and code comprise a body of evidence that can be further used to probe new theories on "classic" questions. I view projects like the Metaketas as a one argument to many studies research design. In contrast, the approach pursued by this paper is a many (several) arguments/theories to one study type design. I see work to better understand the properties of both approaches as a ripe agenda for further work.

Part of this reflection on the merits of collecting original data versus conducting secondary analysis of existing data is purely pragmatic. I write this Introduction during the COVID-19 pandemic, which will likely curtail many field-based projects for at least the next year. This represents a challenge to most field-based research agendas (Wood et al., 2020). As such, the present represents an opportune time to think creatively about how we can learn more from data that has already been collected.

1.4 Takeaways and Future Research

I see two central takeaways from this work, which I intend to develop in future research.

First, I argue that the study of policy implementation merits a more prominent role in the study of "who gets what" than it currently occupies. By increasing the salience of implementation, we open opportunities to learn more about what bureaucrats do and how bureaucracies are organized. This dissertation posits two paths to advance this agenda. First, we need better measurement of the distortions that occur in the mapping of budgets into public goods and services outputs. Chapter 2 provides one approach to this problem focused on measuring bureaucratic behavior directly. Examining more closely the correspondence between allocations and outputs represents a complementary approach (Williams, 2017). Second, better understanding of these processes opens questions about how bureaucrats' efforts to implement policy influence the behavior of other actors to shape broader political outcomes. For example, in Chapter 3, I examine how politicians design oversight in anticipation of these actions, and the consequences of these oversight institutions for distribution. In Chapter 4, I consider how variation in bureaucrats' ability to produce public goods alters a politician's budget allocation strategies and, consequently, a voter's ability to learn and select competent politicians. Understanding variation in bureaucratic implementation processes in the context of broader equilibria in this way widens the potential scope of this research agenda substantially.

Second, I see the call for symmetric study of political phenomena in different contexts as critical to, and often underemphasized in the study of Comparative Politics. The approach that I take toward this mode of investigation relies on the combination of theory and empirics. Data collection or collation is often quite costly, which understandably limits the number of cases in which we could reasonably expect to measure a (micro-level) phenomenon. As such, the empirical results from a single case, as in the Colombian audit experiment in Chapter 2, provide limited leverage on whether we would expect similar effects in other contexts. In my reading, external validity proponents ask whether similar effect sizes (and/or directional effects) obtain elsewhere. My argument for symmetry, instead, seeks to explain similar *or different* outcomes of the same causal process in different contexts.

To this end, the models that I forward present explanations for heterogeneity using models intended to be commonly applied across contexts. In other words, I put minimal scope conditions on the models.[8] Variation in primitives that we expect to vary by context drive variation in outcomes. Paper 2 generates predictions about the direction and magnitude of bureaucratic biases based in service provision on the basis of citizen costs of complaint; politician tastes over constituents; and bureaucratic tastes. Paper 3 generates predictions about the variation in both the adoption and consequences of bureaucratic oversight from variation in bureaucratic quality (selection), bureaucratic insulation, and a politician's preferred constituency in the electorate. Paper 4 explains patterns of political selection and politician malfeasance based on bureaucratic quality and the pool of candidates.[9]

[8]For example, the model of electoral accountability pertains to democracies.

[9]The latter is largely left unexplored in the chapter but is ripe for further investigation.

To my taste, the value of the "comparative" in Comparative Politics comes from the ability to describe and understand both similarities and differences across and within countries. Too often, however, we fixate on similarity given the empirical limitations inherent to the study of difference. The approach that I put forth can be summarized in terms of how scope conditions are invoked. Placing scope conditions on empirical findings (i.e., a directional result) seems unavoidable in many cases. However, we too often use a similar logic to put scope conditions on arguments or theories when they are not necessary. When this occurs, we unnecessarily avoid weighing into differences across contexts. The combination of theory and empirics enables a more precise statement of scope conditions on an empirical finding and also provides some guidance as to how different we could expect outcomes of a common process to be. When we over-scope theory, we are risk being left with a patchwork of arguments from different contexts that complicate the work of understanding political phenomena.

Chapter 2

Bureaucrats Driving Inequality in Access: Evidence from Colombia

Inequalities in access to public goods and services have long challenged developing democracies, inhibiting efforts to reduce socioeconomic inequality and promote economic development. Some inequality is the deliberate result of politicians' budget policies or targeting, but inequalities may also emerge in the production of these goods and services. As "producers of public goods," bureaucrats map politicians' budget allocations into outputs. I argue that bureaucrats are prone to exert differential effort in providing service to different groups of citizens. Variation in bureaucratic effort in turn creates inequality in citizen access to public services. These disparities can emerge even when budget allocations are equitable.

I focus on the setting of service provision, and specifically on interactions between street-level bureaucrats and citizens.[1] In delivering public services to citizens, bureaucrats' actions have distributional consequences. Citizens engage the bureaucracy to gain access to state benefits, from permits to subsidies. While the mode of interaction varies across jurisdictions and services, bureaucrats' role in providing service does not. Critically, bureaucrats distribute public goods and

[1]Following Lipsky (2010), I define street-level bureaucrats as those individuals that interface directly with citizens to implement policies that they do not create.

services because political principals – elected politicians – delegate program administration to bureaucrats. With this delegation comes oversight. Politicians monitor the work of bureaucrats, doling out punishments – from admonishments to termination – and rewards – including bonuses (in some contexts) and promotion – upon observation of bureaucrats' performance.

While politicians ultimately oversee bureaucrats, I consider the role of citizen complaints in directing politicians' oversight. Such complaints to political principals represent the primary form of citizen control over bureaucrats. Complaints function as a means of control by incentivizing the politician to target monitoring to specific decisions of the bureaucrat. Empirically, laws regarding citizen complaints and responses are particularly common in developing democracies and enshrine this form of citizen control over bureaucrats as a right. This paper focuses on differences in citizen propensity to complain as a driver of unequal treatment by bureaucrats that generates inequality in access to services.

This paper makes three innovations. First, I advance a theory that variation in citizen ability to check the bureaucracy via complaint induces bias in bureaucratic effort. The theory emphasizes how such biases map onto inequality in service provision across groups in society. Second, I develop and implement a research design capable of measuring variation in bureaucratic effort behaviorally. In so doing, I generate original measures of effort which allow me to understand how bureaucrats' actions map into ultimate service provision. Finally, I leverage original administrative data to test the bias mechanisms implied by the theory in order to identify the conditions under which bias emerges.

A stylized formal model of service provision underpins my argument about how bias in bureaucratic effort leads to inequality in access to public services. The model demonstrates three mechanisms underpinning bias, which is defined as a difference in the average treatment of citizens from different groups. Following Prendergast (2003, 2007), I argue that citizens' direct mechanism of control over bureaucrats is the complaint to the bureaucrat's principal, here an elected politician. Departing from existing work, I posit that the costs of engaging the bureaucracy through complaints can vary substantially across groups in the population. Where politicians exercise oversight on the basis of such complaints, a rational bureaucrat should exert more effort to provide service to those most likely to complain effectively, inducing statistical bias in effort (Becker, 1957; Phelps, 1972). I refer to this form of bias as *complaint-driven bias*.[2] The model incorporates this complaint-driven bias alongside the bureaucrat's and politician's tastes for providing service to a citizen, the remaining two sources of bias that generate alternate predictions about aggregate levels of bias. I identify that two of three types of bias – the complaint-driven and politician's taste-driven biases – occur precisely *because* there is oversight of bureaucrats.

I measure bias in effort with novel measures of bureaucrats' behavior in a preregistered national-scale phone audit experiment in Colombia. I audited two national social welfare programs in consultation with three national agencies overseeing the bureaucracy and these programs. The audited programs, a conditional cash transfer program (CCT) and a means testing service, have stakes: the CCT program alone confers benefits estimated at 13-17% of household consumption in the median recipient household (Fiszbein et al., 2009). While both programs are funded and directed by the national government, bureaucrats within each local government (*alcaldía*) assume some responsibility for local administration of the programs.[3] Figure 2.1 suggests dramatic variation in municipal administration of the means-testing service (SISBÉN) across Colombia's 1102 municipalities by comparing rates of poverty to rates of program enrollment. Above the horizontal line, SISBÉN has more enrollees than the ostensible municipal population (32.8%). Below the 45-degree line, the program fails to cover the share of the municipal population in poverty – the minimal intended

---

[2] Specifically, this bias is induced by different probabilities of citizen complaint.

[3] For the rest of the paper, I use the Spanish word for local government, *alcaldía*, to refer to the government entities that I study.

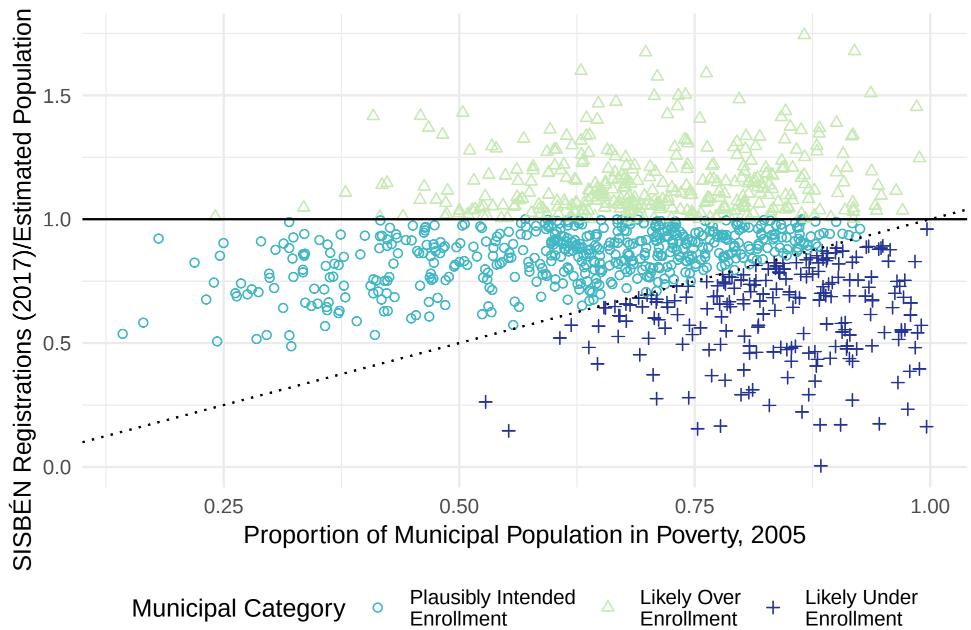


Figure 2.1: The relationship between municipal poverty rates and SISBÉN (a means-testing service) enrollment as a proportion of the population.

population of enrollees and targets of associated social programs (19.1%). Finally, between the two lines, SISBÉN enrollment is plausibly aligned with intended administration (48.1%).

I seek to understand whether bias of street-level bureaucrats in *alcaldías* promulgates such disparities in access. To do so, I direct petitions for information about each service to bureaucrats in *alcaldías* across Colombia. The experiment employs a factorial design that varies both characteristics of petitioners (socioeconomic class, regional accent, and internal migrant status) and the difficulty of the petitions. The experimental design identifies bureaucratic bias in effort. The use of a phone audit offers rich measurement of bureaucrats' behavior, capturing access to officials and provision of information (the service). I then leverage the national scale of the experiment to study the conditions under which bias in bureaucratic effort emerges. This allows me to probe the new mechanism that I propose – complaint-driven bias – and rule out two alternative explanations for bias: taste-based bias and screening. I do this by investigating experimental variation in the difficulty of the petitions and non-experimental variation in welfare program characteristics to understand sensitivity of bias to oversight. I also draw upon an original dataset of all civil servants and contractors in Colombia to measure bureaucrats' incentives as well as demographic and program data characterizing local markets for social welfare services.

I find robust evidence of bias in effort: lower class individuals and internal migrants received substantially less information than their lower-middle class and resident counterparts, respectively. In the context of the audit experiment, I identify these biases on outcomes measuring bureaucratic effort (bureaucrats' actions). I then show that bias in the provision of information in the experiment occurs only in the municipalities where SISBÉN (the means testing program) is administratively underprovided. This finding provides suggestive evidence linking these behavioral outcomes with the ultimate service provision outcomes depicted in Figure 2.1.

Drawing upon the testable implications of the model, I provide clear evidence that the results reflect complaint-driven bias. First, to separate bureaucratic tastes from oversight, I show that bias is attenuated to zero as oversight by politicians becomes becomes less likely. This suggests bias is unlikely to be driven by the bureaucrats' tastes alone. Second, I separate complaint-based bias from politician tastes by showing that anti-poor bias emerges only in poorer places where the ostensible differences between experimental petitioners' abilities to complain is greatest. Finally, I show that the results are inconsistent with screening, an alternate explanation for differential treatment distinct from "bias." Some theories of misallocation of public services suggest that differences in treatment could reflect efforts to screen intended from unintended potential program beneficiaries. However, in contrast to these screening explanations, observed bias in effort cuts *against* the target population for these programs.

Theoretically, this paper unites and extends two foundational approaches to the analysis of agency problems in public service provision. First, it brings the moral hazard problem facing bureaucrats into a large literature on distributive politics focused on relationships between politicians and citizens (voters) (Dixit and Londregan, 1995; Lindbeck and Weibull, 1987). In so doing, I argue that the strategic behavior of intermediary bureaucrats distorts politicians' desired distribution of public services. My findings imply that even if politicians were to allocate funds equally across citizens or constituencies, the lower probability that certain citizens would complain to politicians induces bureaucrats to allocate services unequally.

Second, it brings citizens into models of politician's oversight of bureaucrats (e.g., Banks, 1989; Ting, 2017). In this regard, I endogenize the information that politicians receive to guide oversight by emphasizing the role of citizen complaints to a principal (Prendergast, 2003, 2007). While citizen complaints to a political principal increase service provision, they generate inequality when some citizens are less able to engage politicians via complaint, inducing biased oversight by the principal. I find that oversight can either increase or reduce bias in bureaucratic effort, countering existing assumptions that oversight always deters bureaucratic bias (Hemker and Rink, 2017; White, Nathan, and Faller, 2015).

The empirical findings generate new insights about the administration of social programs in Latin America. In the past two decades, Latin American social policy has expanded to cover well over 100 million former "outsiders," generally low- and lower-middle income individuals (Garay, 2016; De la O, 2015; Huber and Stephens, 2012). These programs – from CCTs to health insurance – are more administratively demanding to implement than past social policies in the region. I join a literature that identifies substantial variation in access to social programs even among ostensible target populations (Holland, 2015, 2017; Niedzwiecki, 2018). In contrast to existing electoral and partisan explanations for this variation, I identify a complementary mechanism through which everyday service provision by bureaucrats generates disparities in access. I provide support for the mechanism through the first experimental audit study of street-level bureaucrats in Latin America.

Taken together, the theory and empirics suggest that socioeconomic inequalities generate political inequalities in citizens' ability to extract oversight over bureaucrats responsible for service provision. These inequalities in voice engender inequality in access to poverty reduction programs intended to mitigate existing disparities. This analysis thus reveals a new mechanism for understanding how inequalities in political voice map onto inequalities in policy outcomes, rooted in seemingly benign everyday interactions between citizens and bureaucrats. By showing that the administration of social programs by bureaucrats can reinforce inequality traps in developing contexts, I highlight the magnitude of the challenge in the design of large scale programs to effectively reduce poverty in light of recent proliferation of these programs in Latin America (De la O, 2015; Garay, 2016).

2.1 Theory

2.1.1 Model

The model consists of three actors: a citizen (or client), a street-level bureaucrat, and a politician, indexed by C, B, and P, respectively. Citizens are differentiated into two groups, $g \in \{x, y\}$ on the basis of observable ascriptive characteristics. Citizens vary in perceived costs to accessing the state. These costs are some function of physical distance, familiarity with bureaucratic procedures, and education.[4] Costs, c_C, are common knowledge and drawn from the distribution C_g which is indexed by group. F_g and f_g denote the cdf and density of C_g, respectively, and $F_g(0) = 0$. Without loss of generality, assume that $F_y(c_C) \leq F_x(c_C)$, or that F_y first order stochastically dominates F_x. Both the bureaucrat and the politician, indexed by i, may have some bias toward providing the citizen of group g with the service. These tastes are represented as $\gamma_i^g \in [0, 1]$, realizations of the random variables Γ_i^g. This bias is strictly taste-based (Becker, 1957). Alternatively, one could conceive of

[4]Rizzo (2018) argues that these barriers are largely psychological; this interpretation is also consistent with my argument.

an altruistic bureaucrat and politician that internalize the benefits when the citizen receives service, with weights proportional to γ_i^g.

A bureaucrat responds to an exogenous citizen request for service by allocating effort, $e \in [0, 1]$. The service is provided with probability e. Effort is costly and is proportional to the difficulty of the task, $c_B \geq 1$. The citizen observes whether she received the service. She subsequently decides whether to complain to the politician, $q \in \{0, 1\}$ at cost c_C. Thus, the costs of complaining are non-trivial and vary across the population as a function of access to the state.

In the subgame in which the service is not realized (provided), the politician receives or does not receive a complaint from a citizen and subsequently chooses a level of effort to invest in auditing the work of the bureaucrat, $a \in [0, 1]$. Politicians audit underprovision as opposed to misallocation of services. This setting characterizes many service provision settings where all citizens have a right to request service. In this sense, the model speaks clearly to the majority of tasks or programs in which equal treatment is a mandate or objective.[5]

With probability a, the politician is able to deliver the service to the citizen. The politician benefits reputationally and thus electorally from the increase in service provision when she detects underprovision, parameterized as $S > 0$.[6] A biased politician will also gain utility from providing the service to a favored citizen. Failing to remedy a complaint induces a separate reputational cost of q. Finally audits are costly, which constrains the intensity of auditing; the marginal cost of an audit on a given task is c_P. To avoid corner solutions, I assume that $c_P > S + 2$. In addition, $c_P > c_B$ implying that it is costlier for politicians to recover the service than for bureaucrats to provide it in the first place. This assumption is consistent with standard arguments about bureaucratic expertise.

[5]Even in the case of social programs with rigorously defined target populations, the first stages of enrollment are generally open to all. If the initial application for, e.g. food stamps were inaccessible, it would be impossible to ensure the program is reaching the entire target population.

[6]In the interpretation of this model with an altruistic politcian, one could consider S as the politician's internalization weight on the service provided to any citizen.

The politician's expected utility can thus be expressed as:

$$E[U_P(a)] = a(S + \gamma_P^g) - (1-a)(q) - \frac{c_P a^2}{2} \tag{2.1}$$

The citizen receives a utility of $b > 0$ if she receives the service. The citizen's expected utility conditional on not having received the service from the bureaucrat is a function of the probability that the oversight process process will recover the service and her decision to complain (q), as expressed in Equation 2.2.

$$E[U_C(q)] = ba - qc_C \tag{2.2}$$

Finally, consider the bureaucrat's utility.[7] He gains utility proportional to γ_B^g by (directly) providing a favored citizen with service. If a decision is reversed during the course of an audit, bureaucrats incur a penalty of $r \in [0,1]$. In practice, these costs range from a reprimand, to transfer, or even termination. The bureaucrat's expected utility is thus:

$$E[U_B(e)] = e\gamma_B^g + (1-e)(-ra) - \frac{c_B e^2}{2} \tag{2.3}$$

2.1.1.1 Sequence

The game proceeds as follows:

1. The bureaucrat chooses an effort level e to provide the service to the citizen.

2. The citizen decides whether or not to lodge a complaint to the politician.

3. The politician chooses the intensity of audits, a. With probability a she overturns the bureaucrat's decision.

[7] One can assume that the bureaucrat receives a fixed wage that satisfies his participation constraint; importantly, in this public sector setting, the wage does not depend on the effort exerted.

4. Payoffs are realized.

I characterize the unique subgame perfect Nash equilibrium (SPNE) in pure strategies. The bureaucrat's allocation strategy sets $e \in [0, 1]$. The citizen's complaint strategy maps the realization of the service provided into a binary decision whether to complain to the politician $q : \{0, 1\} \to \{0, 1\}$. The politician's audit strategy then maps receipt of a complaint into auditing intensity, $a : \{0, 1\} \times \{0, 1\} \to [0, 1]$.

2.1.2 Results

The main results characterize equilibrium effort, which allows for derivation of levels of bias. I solve the model by backward induction, beginning with the politician's decision whether or not to audit the bureaucrat's allocation. The politician's objective is clearly concave in a; differentiating Equation (1) with respect to a yields an interior optimal audit intensity of:

$$a^* = \frac{S + \gamma_P^g + q}{c_P} \tag{2.4}$$

Note that a^* includes two types of oversight. S and γ_P^g represent "police patrols" for underprovision of the service while a complaint, q, represents a "fire alarm" (McCubbins and Schwartz, 1984). The optimal audit intensity allows for analysis of the citizen's optimal complaint strategy. In the subgame in which service is not provided, citizens will lodge a complaint if:

$$\frac{(S + \gamma_P^g + 1)}{c_P} b - c_C > \frac{S + \gamma_P^g}{c_P} b \tag{2.5}$$

This yields an optimal complaint strategy of:

$$q^* = \begin{cases} 1 & \text{if } c_C < \frac{b}{c_P} \\ 0 & \text{if } c_C \geq \frac{b}{c_P} \end{cases} \tag{2.6}$$

This implies that for higher c_C, citizens are effectively "frozen out" of contesting the bureaucrat's service provision. The audit and complaint strategies map directly into the bureaucrat's initial decision on whether to exert effort. Substituting equations 2.4 and 2.6 into the bureaucrat's objective and maximizing, the bureaucrat's optimal effort is:

$$e_g^* = \min\left\{\frac{\gamma_B^g}{c_B} + \frac{r}{c_B c_P}\left(S + \gamma_P^g + \mathbb{I}\left[c_C < \frac{b}{c_P}\right]\right), 1\right\} \tag{2.7}$$

where $\mathbb{I}$ is an indicator function.

Collectively e_g^*, q^*, and a^* characterize the SPNE of the game. In Appendix A.1.3, I endogenize the citizen request for service by assuming that citizens pay a cost proportional to c_C to request the service in the first place. Including this cost introduces two mechanisms through which service provision changes from the baseline results. Clearly, if costs are sufficiently large relative to the benefits of receiving the service, some citizens opt out, receiving no service. Less obviously, it changes the composition of the portion of a group that requests service. This increase in the conditional probability that a citizen that requests service will complain increases the expectation of equilibrium effort across the population.

2.1.3 Defining and Measuring Bureaucratic Bias

There are two measures of bias implied by the model: bias in effort and inequality in outputs, defined in Definition 2.1. Bias in effort corresponds to different equilibrium levels of effort across groups. Inequality in outputs corresponds to different levels of ultimate service provision by group (at the conclusion of the game). I derive these quantities formally in Appendix A.1.2. I assume that the effort and service afforded to each citizen is independent of the effort and service afforded to other citizens. In the context of service provision, if citizens request services at different times or different days, this assumption is plausible. Even in environments in which bureaucrats face unmanageable caseloads such that more effort for one citizen implies less effort for another, so long as citizens receive service on a first-come-first-served basis and the order of petitions is independent of group membership, bias at the aggregate level can be captured by treating cases independently.

Definition 2.1. *Bias in effort. Bias in effort refers to the difference in expectation of equilibrium effort devoted to a citizen from each group, formally,* $\Delta = \mathbb{E}[e_x^*] - \mathbb{E}[e_y^*]$.

For the purposes of characterizing bias empirically or considering the distributional implications of bias, it is useful to define bias between groups in the aggregate. I focus on the case in which effort is interior, i.e. $e_g^* < 1$ for all citizens. I characterize bias in terms of aggregate differences by group. Define differences in the expectation of bureaucrat's tastes as $\eta_B = \mathbb{E}[\gamma_B^x] - \mathbb{E}[\gamma_B^y]$; differences in the expectation of politician's tastes as $\eta_P = \mathbb{E}[\gamma_P^x] - \mathbb{E}[\gamma_P^y]$; and differences in the probability of complaint as $\eta_Q = F_x(c_C) - F_y(c_C)$.

Proposition 2.1. *Between-group bias in effort. The aggregate level of bias between groups* x *and* y *evaluates to:*

$$\Delta = \underbrace{\frac{\mathbb{E}[\gamma_B^x] - \mathbb{E}[\gamma_B^y]}{c_B}}_{\text{Bureaucrat's Tastes}} + \frac{r}{c_B c_P}\left[\underbrace{\mathbb{E}[\gamma_P^x] - \mathbb{E}[\gamma_P^y]}_{\text{Politician's Tastes}} + \underbrace{F_x\left(\frac{b}{c_P}\right) - F_y\left(\frac{b}{c_P}\right)}_{\text{Complaint-Driven}}\right] = \frac{\eta_B}{c_B} + \frac{r(\eta_P + \eta_Q)}{c_B c_P} \tag{2.8}$$

(Proof in appendix.)

Bias in effort between groups x and y can be decomposed into bias that enters through the probability of oversight, Δ_O and bias from the bureaucrat's tastes, Δ_B:

$$\Delta_O = \frac{r(\eta_P + \eta_Q)}{c_P c_B} \tag{2.9}$$

$$\Delta_B = \frac{\eta_B}{c_B} \tag{2.10}$$

Proposition 2.1 implies three mechanisms that drive the bias in effort and outcomes. The differences η_P and η_B capture taste-driven biases of the politician and bureaucrat, respectively. The model also implies the potential for complaint-driven bias, a form of statistical bias, parameterized as η_Q. Note that, in contrast to standard models of statistical bias in which group membership is observable and correlates with some latent trait, I show that this bias emerges even with complete information. When one group is more able to complain, bureaucrats anticipate the increased probability of oversight by giving better service ex-ante. This is captured through a comparison of the distribution of costs for each group. The stochastic dominance assumption serves as a sufficient condition for complaint-driven bias to emerge on average (in the aggregate).

Of the three sources of bias, the complaint-based bias and the politician's taste-based bias are driven by oversight of the bureaucrat by the politician. In this sense, both forms of bias are *strategic*. Oversight is biased if $\eta_P + \eta_Q \neq 0$. This implies that the politician exerts more effort in auditing members of one group than another when service is not provided by the bureaucrat. In focusing on bureaucratic behavior, I emphasize that the probability of audit conditions the bureaucrat's level of effort. Without loss of generality, assume that oversight is biased in favor of group x, e.g. $\eta_P + \eta_Q > 0$.

Proposition 2.2. *Bias and the likelihood of oversight. Given $\eta_P + \eta_Q > 0$, a higher probability of audits for citizens of group x citizens increases the magnitude of the bureaucrat's bias in effort if and only if $\frac{r(\eta_P+\eta_Q)}{2c_P} > -\eta_B$. The higher probability of audits for citizens of group x will only reduce the magnitude of bias in effort if $\frac{r(\eta_P+\eta_Q)}{2c_P} < -\eta_B$. (Proof in appendix.)*

Thus, oversight can increase or decrease the level of bias in effort exerted by bureaucrats. Critically, in order for oversight to decrease bias, if oversight-driven biases favor group x, the bureaucrat's tastes must favor y ($\eta_B < 0$) and be sufficiently large in magnitude. This result emerges because the politician optimizes service provision (possibly with some preference to one group), not equality in access. I extend this analysis to investigate the relationship between oversight and bias in outputs in Appendix A.1.4. Importantly, I find that if oversight is biased, bias in effort is a sufficient condition to generate inequality in outputs.

2.1.4 Testable Implications

The model posits three types of bureaucratic bias in effort. These biases emerge in the bureaucrat's original decision to devote effort to provide service. The distributional consequences of bias in effort for "who (ultimately) gets what" services depend on what is driving these biases. For this reason, it is important to disentangle the mechanisms underlying any observed patterns of bias. While the model implies no direct econometric test of the mechanism, Proposition 2.3 derives several testable implications that I use to discriminate between types of bureaucratic bias.

Proposition 2.3. *Tests of the mechanism. Decomposing oversight-driven and non-oversight-driven bias:*

1. *Bias in effort varies in the politician's cost of auditing if and only if oversight is biased:* $\frac{\partial \Delta}{\partial c_P} \neq 0$, *if and only if* $\Delta_O \neq 0$. *For sufficient increases in* c_P, *bias in effort attenuates toward zero.*

2. *The magnitude of bias in effort increases in the strength of bureaucratic incentives if and only if oversight is biased:* $\frac{\partial \Delta}{\partial r} > 0$ (< 0), *if and only if* $\Delta_O > 0$ (< 0).

Discriminating politician's tastes from complaint-driven bias:

1. *The magnitude of bias in effort increases in the between-group differences in ability to complain:* $\frac{\partial \Delta}{\partial \eta_Q} > 0$ (< 0) *if* $\eta_Q > 0$ (< 0).

	Classification			Testable Implications			
Bias Mechanism	Bias Type	Oversight-driven	Case	$\frac{\partial\Delta}{\partial c_P}$	$\frac{\partial\Delta}{\partial r}$	$\frac{\partial\Delta}{\partial\eta_Q}$	$\frac{\partial\Delta}{\partial\eta_P}$
Complaint-driven: Citizens from group x are more likely to complain than from group y which draws a higher likelihood of auditing by the politician. The bureaucrat devotes more effort to x in anticipation of higher probability of audit.	Statistical	Yes	$\Delta > 0$	$\neq 0^*$	> 0	> 0	0
			$\Delta < 0$	$\neq 0^*$	< 0	< 0	0
Politician's tastes: The politician prefers to audit service to group x more than to group y. The bureaucrat devotes more effort to x in anticipation of higher probability of audit.	Taste-based	Yes	$\Delta > 0$	$\neq 0^*$	> 0	0	> 0
			$\Delta < 0$	$\neq 0^*$	< 0	0	< 0
Bureaucrat's tastes: Bureaucrat prefers providing service to group x over group y.	Taste-based	No	$\Delta > 0$	0	0	0	0
			$\Delta < 0$	0	0	0	0

Table 2.1: Summary of the bias mechanisms implied by the theory and the testable implications for distinguishing the mechanisms. Bias is defined as a difference between two groups: when $\Delta > 0$ x is preferred to y and when $\Delta < 0$, y is preferred to x, so testable implications should be seen as *magnitudes*. *Note that sufficient increases in c_P attenuate bias toward zero.

2. *The magnitude of bias in effort increases in the between-group differences in the politician's tastes:* $\frac{\partial\Delta}{\partial\eta_P} > 0\ (< 0)$ *if* $\eta_P > 0\ (< 0)$.

Proposition 2.3 guides efforts to test the mechanisms described in the model and in Table 2.1. I proceed in two steps. First, I test for evidence of oversight-driven bias. This distinguishes bias coming from bureaucrats' tastes (Δ_B) from the bias that comes from different probabilities of oversight (Δ_O). As in Table 2.1, oversight-driven bias incorporates both politician tastes and complaint-driven bias. To do this, I examine variation in bias with respect to the two parameters that should only drive variation in the oversight mechanism, the politician's cost of effort, c_P, and the "bite" of possible punishment, r. If bias varies in these two measures, there is evidence of oversight-driven bias.

Conditional on finding evidence of oversight-driven bias, I aim to distinguish between politician tastes and citizen propensity to complain, the two components of oversight-driven bias. To do this, I examine variation in citizens' cost of complaint (enters through η_Q). As the "distance" between petitioner types' costs of complaint, η_Q, increases, so too should the relative magnitude of complaint-driven bias. In particular, if the magnitude of bias increases as η_Q increases, there is evidence that bias comprises complaint-driven bias. This test is able to distinguish between politician tastes and complaint-driven bias when $Cov(\eta_P, \eta_Q)$ is small. I also consider a parallel test with regard to politician incentives that may drive politician tastes, η_P with a parallel logic.

2.2 Case Context

I measure variation in bureaucratic discretion at national scale in Colombia. Writings on state capacity in Colombia have long focused on a two-century history of civil wars as a cause or consequence of state weakness (Centeno, 2003; González González, 2014). Nevertheless, studies of bureaucracy typically characterize Colombia's national bureaucracy as comparatively "Weberian" by regional standards (Evans and Rauch, 1999; Mayka, 2016). Analyses of the World Bank's Worldwide Governance Indicators echo these findings (see Appendix A.2.2). Aside from measures of political violence, the other governance indicators approximate the world median, rank in the highest tercile in Latin America, and rank in the lowest decile of OECD countries.[8] At the national level, bureaucratic capacity in Colombia is believed to vary with the relative concentration of "technocrats" and patronage employees (e.g., Schmidt, 1974; Dargent, 2016). Less is known about municipal bureaucrats, the subjects of this investigation. However, dozens of semi-structured interviews with national bureaucrats and participant observation in *alcaldías* evidence tremendous variance in professionalism, competence, and outputs.

[8]Colombia joined the OECD in July 2018.

2.2.1 Municipal Politics in Colombia

Since political decentralization in the late 1980s, Colombia's 1102 municipalities have assumed responsibility for most services, ranging from roads to education.[9] Decentralization gave rise to larger and relatively more professional municipal public administration (Fizbein, 1997). Important for this project, some national programs are implemented "on the ground" within municipalities by municipal bureaucrats.

Municipalities are governed by a mayor and local council (*concejo*) of seven to 45 councilors, according to population.[10] Mayors are elected every four years by plurality vote and are limited to serving one consecutive term. In contrast, councilors are elected by optional open list PR without term limits. In these elections, parties are weak and the role of ideology in elections is limited. According to the Colombian party classification by Fergusson et al. (2018), 232 current mayors represent parties classifiable as "left" or "right;" 543 mayors represent parties without an identifiable ideology; and 325 mayors ran without a party (as their own party). In these contexts, the distribution of public and private goods arguably constitutes the basis of political competition.

2.2.2 Bureaucratic Hiring

At the municipal level, bureaucrats are hired and overseen by local politicians. Politicians staff the public sector via two hiring mechanisms: civil service (called *empleados de planta*) and contracting. The degree to which mayors delegate staffing the *alcaldía* varies predictably with the size of the municipality. In small municipalities, contracts are signed directly by the mayor (on behalf of the *alcaldía*); in larger municipalities, high-level political appointees sign off.

On average, contractors are less expensive to employ than civil servants. From a purely practical

[9] There are 1122 territorial units, however 20 are *corregimientos* as opposed to municipalities. There were elections in 1100 municipalities 2015; it is from these municipalities that the sample of experimental municipalities was drawn.

[10] Bogotá has a local council of 45 councilors; the next largest councils have 21 members.

standpoint, contractors have higher powered incentives than are typical in canonical public sector settings (Dixit, 2002). Contracts are short term, on average less than five months, whereas civil service employees empirically enjoy longer tenure and thus higher job security. The processes of contracting are relatively lax and the share of contractors that could reasonably be considered "patronage hires" is certainly higher than the share of of civil servants, though this is an admittedly imperfect proxy.

Municipalities are legally constrained to a maximum budget share that they can devote to remuneration of civil servants. These constraints are a function of a municipal classification based on revenues and population (*Ley 617 de 2000*). Municipal workforces are thus supplemented by contractors. From the perspective of a politician, contracting provides a relatively flexible means of delivering jobs. Yet, limited tenure provides few opportunities to develop expertise and, as practiced, contracting yields high fluctuation in actual staffing levels within the *alcaldías* throughout the year, possibly reducing productivity of the bureaucracy.

2.2.3 Complaints and Oversight

I posit a fundamental role for citizen complaints as a means of seeking oversight over bureaucrats. As in much of Latin America, Colombia provides substantial legal rights for making complaints. The Colombian Constitution of 1991 mandates the right to access public information for all citizens (Article 74) and statutory law allows Colombians to request "recognition of rights, intervention of a government entity or official, legal resolution, service provision, information, copies of documents, consultations, [various forms of] complaints, and claims" by written petition (translated from *Ley 1755 de 2015*, Article 14). In turn, the government has three weeks (fifteen business days) to respond to the petition. Complaints also emerge through less formal channels; the distinction is not relevant for the purposes of the theory so long as costs vary across the population.

Some complaints are ostensibly handled by other higher-level bureaucrats, while others rise to local mayors. While such instances are, in principle, hardly newsworthy, mayors do audit the local administration of social programs. Yet even in this context, there are regular news reports of mayors responding to complaints about the function of social programs, typically by auditing local rolls of beneficiaries.[11] For such action to influence bureaucrats' behavior, threats of reprimand must be perceived. An original survey of street-level bureaucrats in *alcaldías* in Bogotá and Cundinamarca finds that 78% (57/73) of these bureaucrats perceive that a mistaken decision would be punished (with varying severity) and decisions would be reversed (see Appendix A.3 for survey information).

While nationwide data on complaints to local entities is not collected, analysis of over 440,000 complaints filed in public entities in Bogotá from January 2017 through June 2018 and compiled by the Veeduría Distrital provides two stylized facts of note. First, virtually all complaints relate to service provision and approximately 125,000 complaints (28 percent) explicitly relate to bureaucrats' actions in service provision. This represents the most common class of complaints. Second, leveraging Bogotá's geographic segregation by class, analysis of complaints at the locality level suggests that there are substantially *more* complaints per capita in richer localities. Figure A.4 shows a positive association between the the average class designation (*estrato*) of residential properties in each locality and the per-capita rate of complaint submission to each local *alcaldía*. This pattern emerges despite the fact that service provision is notably better in richer localities – those where complaints are most frequent.[12] (See Appendix A.4 for details.)

[11]Recent newsworthy investigations include investigation of how a councilor in Mosquera, Cundinamarca made it onto a list of means-tested beneficiaries for social programs (SISBÉN); a scam to stuff the rolls for Adultos Mayor, a subsidy for senior citizens, in Florencia, Caquetá; and a general audit of the SISBÉN rolls in Pitalito, Huila. The first two investigations occurred in response to citizen complaints.

[12]The model implies that the group with higher barriers to complaint (the poor) could, in equilibrium, complain more than the rich. If the rich receive better services such that they do not need to complain, the poor could complain at higher rates. To the extent that the rich do complain more frequently despite receiving better service, the model provides qualified evidence against taste-based bias of a high magnitude (see Appendix A.1.5).

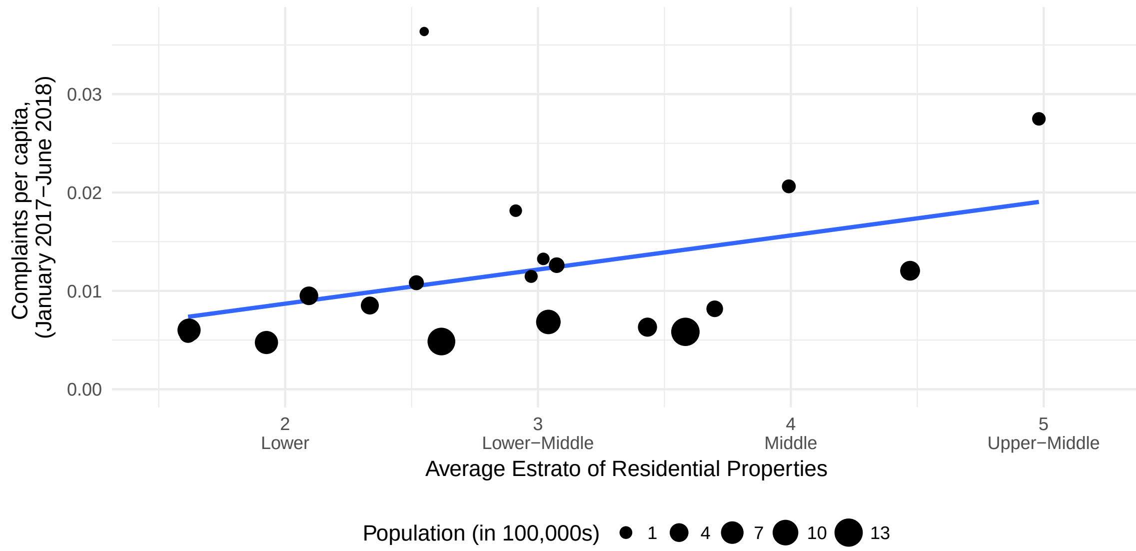


Figure 2.2: Rate of complaints filed by locality in Bogotá by average wealth of the locality.

2.2.4 Audited Social Welfare Services

I audit two nationwide social welfare programs that are administered, in part, by officials embedded in every municipal government. Specifically, the national government agencies that oversee these programs maintain agreements (*convenios*) with each municipality that mandates that local *alcaldías* hire the local program officials. The programs, the System for the Identification of Beneficiaries of Social Programs (SISBÉN) and Más Familias en Acción (MFA), provide coverage on a nationwide geographic scale. These programs provide access to or transfers to low-income Colombians. Colombia is among world's most unequal countries with a very large low-income population, indicating a large but not universal pool of potential beneficiaries for each program. Existing analyses of the programs suggest very different levels of politicization between the two programs.

Created in 1995, SISBÉN is a household index of assets used for qualification for means-tested social programs.[13] SISBÉN is a prerequisite to access subsidized health insurance and most social

[13]SISBÉN is not a social program in the conventional sense; it does not confer a direct benefit on registrants. However, this is the means by which Colombians access social benefits and is widely regarded as a central piece of the social policy

programs, including Más Familias en Acción. The municipal service associated with SISBÉN is the administration and readministration of household survey of assets. The survey is then sent to the National Department of Planning (DNP), which generates a score on an index through a private formula. At present, over 36 million Colombians are registered in the SISBÉN system, representing approximately 70 percent of the population. SISBÉN is thought to be manipulated by citizens (i.e. by hiding assets during the survey) or local bureaucrats. Past iterations of the formula have been changed to reduce local discretion. I identify many municipalities with far more SISBÉN registrations than the ostensible population, supporting longstanding views of politicization (Camacho and Conover, 2011).

Created in 2002, MFA (formerly *Familias en Acción*), is Colombia's national conditional cash transfer program that provides subsidies to (mostly) mothers on the basis of compliance with their children's health and educational requisites. Each municipality has one MFA official (*enlace*), though the program is overseen by a much larger office in larger municipalities. These individuals provide information to recipients and monitor compliance with community participation aspects of the program. MFA enrolls between 15-20 percent of Colombian households, providing subsidies estimated at approximately 15 percent of median household consumption among recipients (Fiszbein et al., 2009). In contrast to SISBÉN, MFA is designed and implemented as per transparent and uniform programmatic guidelines, at least relative to other conditional cash transfer programs in the region (De la O, 2015).

2.3 Research Design

To test the implications of my theory of bureaucratic biases, the research design must be able to (a) elicit bias and (b) measure bureaucratic effort, the outcome of interest. I utilize phone audits to facilitate direct behavioral measurement of bureaucrats' response to requests for service (informa-

landscape.

tion). I implement these audits at national scale in Colombia. In order to elicit bias, I randomly assign petitioner characteristics. The random assignment permits direct identification of bias, Δ. I also randomize the several characteristics of the petition itself to test theoretical claims about the sensitivity of bias to the cost of effort and to reduce detection of the audits.

2.3.1 Audit Experiment

The unit of random assignment is the petition: a call with a request for a service. I utilize a factorial experiment to randomly assign characteristics of petitions. The treatments provide exogenous manipulations in the bureaucrat's and politician's marginal cost of effort, c_B and c_P, as well as in observable attributes of the petitioner, g. Several aspects of petitions are constant across calls. Given the composition of MFA recipients – mostly mothers – all petitioners are female. While all calls were made from Bogotá, the outgoing phone numbers appeared as standard cell phone numbers. In Colombia, cell phone numbers do not convey geographic information.

2.3.1.1 Treatments

In order to induce experimental variation in bias, I manipulate identity-based characteristics of the petitioners. These characteristics are rooted in the Colombian social and political context, and serve as the analogue to the groups in the theoretical model. First, I assign the socioeconomic class (*estrato*) of the caller. Since independence (and before), class has represented an organizing feature of Colombian society and political life (Martz, 1997; Sanders, 2004). Given the focus of the social programs audited, I differentiate between low- and lower middle-class callers.[14] Focus groups with Colombians of different socioeconomic classes and observation of calls in a government call center suggest several avenues in which the class of a caller can be immediately distinguished by phone. Specifically, callers in the two groups vary in their vocabulary, salutations for figures

[14] For readers familiar with Colombia's class categorization system, lower class refers to *estratos* 1 and 2. Middle class refers to *estrato* 3. Politically-defined class categories (*estratos*) range from 1 to 6. Appendix A.5.2 reflects the distribution of individuals by class in the population as of 2005.

of authority (bureaucrats), and the framing of questions. While class communicates a variety of features, I assume that on average lower middle-class individuals have relatively greater ability to engage (access) the bureaucracy than lower-class individuals.

A second identity-based manipulation relies on regionalism in Colombia. Due to colonial settlement patterns, a rugged topography, and limited central government penetration, social and political life flourished within regions in the 19th century (González González, 2014; Uribe de Hincapié and Álvarez Gaviria, 1998). Two centuries later, regional accents remain quite distinct and regions maintain relatively distinct cultures and political organization (Ocampo, 2014). I randomly assign Bogotano, Paisa, and Costeño regional accents. These accents represent the most widely-spoken accents in Colombia and are collectively spoken by $\approx$60 percent of the Colombian population. Appendix A.5.1 includes maps of the geographical coverage of these accents.

The accents probe concepts of embededness of petitioners. Recent arguments have emphasized the embeddedness of bureaucrats within state or local governments, using bureaucrat region (resp. state) of origin to measure "embeddedness" (Pepinsky, Pierskalla, and Sacks, 2017; Bhavnani and Lee, 2018). Extending this logic, in a context where the vast majority of local bureaucrats are drawn from local communities, regional accent should provide a signal of embeddedness or lack thereof. There is not an obvious ranking of costs of access within the accents, though long-held stereotypes and data on outcomes hold that service is substantially better in the highlands (home of the Bogotano and Paisa accents) than on the Caribbean coast (home of the Costeño accent).

The third identity-based treatment focuses on migrant status. This treatment is communicated during the petition through a statement that the individual in need of the service is a recent (internal) migrant. The "resident" condition does not provide this information. Rates of internal migration have long stood among the highest in Latin America and encompass both ordinary and conflict-induced migration (Martine, 1975; Ibáñez and Vélez, 2008).[15] See Appendix A.5.3 for estimates of rates of internal migration in Colombia. The migrant condition signals two potentially relevant features. First, migrants are apt to have less familiarity with a municipality, suggesting higher costs of access. Second, following Gaikwad and Nellis (2017), migrant status indicates a much lower likelihood of voting, as there is no absentee voting and re-registering in a new municipality is cumbersome and occurs during relatively short temporal windows prior to elections.

The final manipulation varies the technical specificity of the petition.[16] For both programs, the "easy" version of the question simply asks how a non-enrolled/registered citizen could enter SISBÉN or MFA, respectively. The "technical" version of the questions poses a question about a situation with specific technical program requirement. This manipulation allows me to test sensitivity of bias to costs to the bureaucrat which aids in discriminating between bias mechanisms.

Collectively, all four factors are fully crossed, yielding a $2 \times 3 \times 2 \times 2$ factorial design summarized in Table 2.2. This yields 24 distinct treatments for each of two programs, though I analyze along the margins (by attribute). Note that the twelve confederates were actresses. All confederates voiced both low- and middle-class petitions. To maximize authenticity, actresses voiced only their own regional accent and calls were divided between four actresses per region of origin. Calls were randomly assigned to each confederate.

All calls were recorded. I hired Colombian coders to listen to all of the recordings to double code call characteristics and responses. Given that the coders were blinded to treatment assignment, this yields one measure of compliance with treatment assignment. I define compliance as a measure of whether coders reported hearing the assigned factors (i.e. if they heard a Costeña petitioner on a

[15]For example, 7.4 million Colombians are internally displaced, representing approximately 15% of the Colombian population.

[16]Under the assumptions of the model, that $c_P > c_B$ this may or may not also increase the politician's cost of effort c_P commensurately.

Factor	Levels	Mode of Administration	Compliance Rate
EFFORT (COSTS)			
Difficulty of Request	• Easy • Difficult	Technical specificity of request to petition, as defined by national government partners	99.3% 99.2%
BIAS			
Regional Accent	• Bogotá • Paisa • Costeño	Regional accent of caller employed in interaction with bureaucrats.	99.7% 98.4% 98.7%
Socioeconomic Class	• Low • Lower Middle	Vocabulary, salutations, and framing of the interaction.*	76.7% 79.3%
Stated Migrant Status	• Migrant • Resident	One statement in delivery of petition (migrant). No reference to internal migration in resident's call.	97.3% 95.0%

Table 2.2: Factors and levels employed in the factorial design. Compliance rates are calculated as the proportion of calls correctly classified by double coders out of the number of calls assigned to each level for which the factor was revealed (see Section 4.3 for details on the rollout and estimation). *Note that while the framing of the interaction varied across class but the statement of the question itself was stated identically for both classes.

call assigned to a Costeño accent). The rates of compliance are reported in the final column of Table 2.2. A more detailed analysis of compliance is reported in Appendix A.11.1. While I cannot know what bureaucrats intuited, rates of compliance in the double coding exercise are quite high across all factors and levels, alleviating major concerns.

2.3.1.2 Sampling, Assignment

The sample of *alcaldías* was selected with two opposing objectives. First, by maximizing the number of petitions made to the same *alcaldía*, I increase statistical efficiency and allow the estimation of within-*alcaldía* treatment effects. Second, I seek to minimize the probability of detection. In order to achieve both objectives, I stratify municipalities into three groups by estimated 2018 population. Note that Bogotá provides services at the level of 20 localities. The entities are thus municipal *alcaldías* outside of Bogotá and local *alcaldías* in Bogotá. The number of petitions varies by stratum. In the large stratum, six petitions were assigned, three each for SISBEN and MFA. In the medium stratum, four petitions were assigned, two per program. In the small stratum, one petition

				n Petitions per Entity			
Stratum	Stratum Size	Population threshold	Sample	SISBÉN	MFA	Total	Total Petitions
Large	80	> 100,000	All	3	3	6	480
Medium	140	[35,000, 100,000)	All	2	2	4	560
Small	898	< 35,000	398	1	1	2	796
Total	1118		618				1836

Table 2.3: Sample of municipalities (or localities) and number of petitions. Note that in the small stratum, localities are selected proportionally to population size. All population data from 2018 estimates from DANE.

was assigned per program. The distribution and number of petitions is depicted in Table 2.3.

Blocking by *alcaldía* was used in order to ensure maximal within variation and avoid detection. The blocking procedures are detailed in Appendix A.9.1. The blocking ensures that each *alcaldía* received equal numbers of low- and middle-class petitioners; equal numbers of easy and difficult questions; and received half the petitions from migrants. To minimize the likelihood of detection, the more specific technical questions were never repeated within an *alcaldía*. This implies that the ratio of easy to technical questions in the large stratum was 2:1. The estimation strategy accounts for these differential probabilities of assignment. Further, no *alcaldía* received more than one call from the same class/accent combination or was asked the same question more than once.

The order of calls was randomly assigned to space out calls to the same *alcaldía* over approximately four weeks. The assignment process for this rollout procedure is documented in Appendix A.9.3. In general, first attempts of each call were consistent with the assigned ordering (within morning or afternoon), but repeated attempts complicate this mapping. Finally, the time of day – morning or afternoon– within each *alcaldía's* hours of service was randomly assigned. Each *alcaldía* received equal numbers of calls at each time. Ultimately, just 6 calls were detected. Analysis of the detected calls yields no systematic patterns, minimizing concerns in this regard (see Appendix A11.2).

2.3.1.3 Outcomes

The audits measure a rich set of behavioral outcomes relating to service provision through the course of the call. Appendix A.6 clarifies the sequencing of calls and outcome measurement. To measure service provision, all enumerators filled out an instrument to document the trajectory, outcomes, and information conveyed in each call. Further, all calls were recorded. Two trained research assistants listened to every recording and double entered all data, including additional measures of compliance and qualitative observations of each call.

I focus on three classes of outcomes. For the *alcaldías* reached by phone, I provide a mapping of the call through the *alcaldía*. Since dispatchers who answer are not generally program officers, I measure whether a petitioner was provided access to a program officer in order to make the petition. I map the mode of transmission through the bureaucracy to measure the accessibility and navigability of service providers within local bureaucracies. In particular, I measure four outcomes dichotomously: (1) whether the dispatcher identified himself/herself; (2) whether the petitioner was able to make (state) the petition; (3) whether the petitioner was connected to at least a second official; and (4) whether a program officer for SISBÉN or MFA from an ex-ante pre-treatment list was identified.

Most important, I measure agents' responses to the petition. I focus on the amount and veracity of information provided relative to the benchmark (correct) answers specified by the national government agencies that oversee each program. Outcomes at this stage also include a measure of red tape: whether an official asked for *extra* requirements not specified by program guidelines and whether petitioners were asked to come "in person" without further guidance. I measure five pre-registered outcomes of interest: (1) whether the correct, complete answer was provided; (2) whether partial information was provided; (3) whether any actionable information was provided (a sum of #1 and #2); (4) whether the petitioner was asked to come to the *alcaldía* in person without further instruction; and (5) whether red tape was solicited. The "come to the *alcaldía*" response merits some clarification. All services require an eventual trip to the *alcaldía*. Arriving without the requisite documents imposes additional costs on the petitioner, regardless of the bureaucrat's intent.[17]

Finally, I use confederate ratings of service as a benchmark to the behavioral measures of service provision. Here I examine whether the perceptions of the petitioner align with experiences of service. A z-score index includes assessments of competence, knowledge, respect, trustworthiness, and satisfaction.

2.3.1.4 Ethical Considerations

Government audit experiments generally raise three ethical concerns: the use of deception, the protection of subjects, and the waste of time and public resources. I address the concern of deception through a novel model of collaboration with national government agencies. The collaboration included consultation throughout the research design process with the agency overseeing the Colombian bureaucracy at the national level (the Administrative Department of Public Administration) as well as the agencies overseeing SISBÉN (National Department of Planning) and MFA (Department of Social Prosperity). These agencies provided guidance on the programs to be audited, the content of the audits, the correct answers to the audits, and some administrative data. In exchange, I conducted the experiment independently with external funding and produced and presented a policy report to each agency in June 2018.

Notably, these agencies conduct their own "mystery shopper" (*cliente incógnito*) audits of em-

[17]Two plausible interpretations of the "come to the *alcaldía*" response include: (a) political capture is more likely to occur in person than on the phone; or (b) the bureaucrat believes that the petitioner will only understand in person. I remain agnostic between these interpretations but maintain that failure to provide information imposes an additional cost to petitioners.

ployees and contractors periodically, though my collaborators do not recall randomizing any components. By conducting the audits independently, I provide additional privacy protections to subjects (audited bureaucrats) in a manner that cannot be guaranteed in government audits.

In terms of wasting of time and resources, the costs to public entities in Colombia should be weighed against the benefits of this original data and report. The upper bound on the costs to these entities can be quantified quite simply. The answered calls (i.e. those that occupied the time of public employees) total under 200 hours. At the maximum monthly salary for the maximum rank of employee ("*Profesional*") that would have spoken with a caller, the upper bound on the cost of these calls totals $\$2,644$ USD.[18] This totals less than 10 months for one employee at the official minimum wage, a common local benchmark.

2.3.2 Administrative Data

In order to understand the relationship between bureaucratic organization and behavioral measures of service provision by street-level bureaucrats, I leverage several original administrative datasets on public sector personnel in Colombia. The first datasets contains individual public employees working as civil servants in Colombia with self-entered name, position, work experience, and education. Outside of Bogotá, I use the data from the *Sistema de Información y Gestión del Empleo Público* (SIGEP) and inside Bogotá I use the city-level equivalent (SIDEAP). This provides data on public employees hired under the law for public employment.[19]

Second, I generate a list of contractors working for municipal governments using data from Colombia Compra Eficiente, the national government entity that oversees public procurement. This source contains data on public sector contractors working in all government entities. While con-

[18]Calculated from *Decreto No. 309 de 2018.* Maximum public sector salaries are benchmarked by municipal "category," a measure of population and local development. This calculation uses the highest salary in the "special" category of municipality (highest paying) and is thus a strict upper bound.

[19]*Ley 80 de 2003.*

tracting is generally cheaper than hiring civil servants, it concurrently serves as a means to preserve patronage in the face of civil service laws. As such, not all contractors are patronage employees, but my measurement relies on the assumption that contractors are more likely than civil servants to be patronage employees and that aggregate patterns of contracting by *alcaldía* measure the use of contracts for patronage.

Both datasets are entered and maintained by officials within each *alcaldía*. In full, 82 percent of the employees reached in the experiment and program officers appear in this combined list. In the cases in which I am unable to identify the employee, 4.7 percent come from municipalities that do not use one or both datasets.

I also leverage additional demographic and electoral data. Demographic data on the characteristics of municipalities allow me to contextualize the identities portrayed in the calls to local constituencies. In terms of the theoretical model, such data provides some information about the shape of the distribution of costs, f_g, within a given population of citizens.

With the electoral data compiled at Universidad de los Andes, I seek to measure political competition, a feature which should increase the politician's incentives to provide public goods, S, and may covary with the tastes of elected politicians, γ_P^g. Standard measures of political competition are complicated by features of municipal politics in the Colombian context. First, at the municipal level, party labels do not signal ideology and high rates of party switching suggest that analyses at the party level contain little meaningful information. Further, measures based on raw electoral data such as mayoral margin of victory (distance between the winner and runner up) exhibit very little serial correlation.[20] Thus, observing a close election at time t provides essentially no information about competitiveness at $t + 1$. As a result, I develop other measures of political competition. In

[20]The quality of electoral data is generally high. As such, lack of serial correlation is ostensibly driven by characteristics of electoral competition, not data limitations.

particular, I look at the frequency with which individuals are re-elected and which family names concentrate among local council members (*concejales*) over a twenty year panel (six electoral cycles). These measures build upon those used by Acemoglu et al. (2008) to measure political inequality in the department of Cundinamarca, Colombia in the nineteenth century.

2.3.3 Estimation

The estimation of causal quantities in the experiment accounts for the process of selection and the delivery of treatment during the course of the interactions with local government officials. Post-treatment selection represents a threat to inference in existing audit experiments (Coppock, 2018). In the present experiment, if a Costeña petitioner was more likely to be able to state the petition, conditioning the sample on having made a petition may induce bias in estimates of the effect of accent on informational responses.

To overcome this limitation, the attributes (factors) in the factorial design were revealed at three distinct points in the call, as depicted in Table 2.4. This defines three relevant samples: all attempted calls, all answered calls, and all calls in which the petition was delivered. Factors not yet revealed in a given sample are referred to as *placebos*; factors revealed within the sample are referred to as *treatments*; and factors revealed prior to revelation of the a sample are regarded as *pre-treatment covariates*. Point estimates on the treatment variables (in the relevant sample) are causally identified. Taking advantage of the rollout of factors during the course of the call increases statistical efficiency and while avoiding the threat of bias induced by post-treatment sample selection.

I seek to estimate the Average Marginal Component Effect (AMCE) of the randomly-assigned treatments. This effect is the marginal effect of each factor, averaged over the joint distribution over other factors. I account for the differential probabilities of assignment to easy and technical questions across the strata of municipalities with two estimators. I estimate the sample AMCE using inverse probability weighting (IPW) or *alcaldía* fixed effects.[21] The latter strategy examines differential treatment of petitioners leveraging variation only from within the same *alcaldía*.

I estimate the AMCE with either estimator with regressions of the form of Equation 2.11 using OLS with heteroskedasticity-robust standard errors. Note that these standard errors correspond to the level of treatment assignment: the petition. The set of indicators in the regression model corresponds to the factor levels in the design, here $\mathbf{Z} = \{\text{Afternoon}_i, \text{Technical}_i, \text{Lower Middle Class}_i,$ $\text{Bogotá accent}_i, \text{Costeño accent}_i, \text{ and Resident}_i\}$. In Equation 2.11, $\boldsymbol{\psi}_m$ indicates municipality fixed effects; IPW specifications do not include this term.[22] κ_p indicates a vector of program (SISBÉN or MFA) fixed effects that are included in all specifications.

$$Y_{ipm} = \sum_{j \in \mathbf{Z}} \beta_j Z_i^j + \kappa_p + \boldsymbol{\psi}_m + \epsilon_{ipm} \tag{2.11}$$

In order to estimate the conditional AMCE with respect to institutional, demographic, or political covariates, I estimate Equation 2.12 using OLS with heteroskedasticity-robust standard errors. In this equation, moderators and covariates are represented by the variable X_i (resp. X_m). The conditional AMCEs estimated in are causally identified under the conditions specified above for the AMCE. The conditional AMCEs are estimated by β_j and $\beta_j + \gamma_j$, where j indexes the treatment level. The difference in conditional AMCEs (γ_j) is not causally identified absent additional assumptions.

$$Y_{ipm} = \sum_{j \in \mathbf{Z}} \beta_j Z_i^j + \sum_{j \in \mathbf{Z}} \gamma_j Z_i^j X_i + \kappa_p + \sum_{p \in P} \alpha_p X_i + \epsilon_{ipm} \tag{2.12}$$

[21] The implied AMCE estimand coming from the two estimators is subtly different. For the IPW estimator, the sample AMCE (for a factor Z) is $\mathbb{E}[Y_{ipm}(Z_i = 1, \mathbf{Z}) - Y_{ipm}(Z_i = 0, \mathbf{Z})]$, where Z_i is the factor of interest, $\mathbf{Z}$ is a vector of all other attributes. For the FE estimator, the AMCE is given by $\sum_{m \in M} w_m \mathbb{E}[Y_{ipm}(Z_i = 1, \mathbf{Z}) - Y_{ipm}(Z_i = 0, \mathbf{Z})]$, where w_m is a weight proportional to the inverse of the variance within the block.

[22] Fixed effects are highly prognostic of outcomes. Regressions of the main outcomes on a vector of municipality fixed effects yield R^2's of 0.37 to 0.55.

Call Made		$\longrightarrow$	**Call Answered**		$\longrightarrow$	**Petition Made**	
1836 Calls			1194 Calls			911 Calls	
618 Municipalities			466 Municipalities			424 Municipalities	
(Time of Day)	✓		(Time of Day)	Not point identified		(Time of Day)	Not point identified
Accent	Not revealed		Accent	✓		Accent	Not point identified
Class	Not revealed		Class	✓		Class	Not point identified
Difficulty	Not revealed		Difficulty	Not revealed		Difficulty	✓
Migrant Status	Not revealed		Migrant Status	Not revealed		Migrant Status	✓

Table 2.4: Timing of treatment delivery during the process of a call. Attributes that are "not yet revealed" serve as placebos in the outcomes prior to their revelation. The timing of treatment delivery defines the relevant sample upon which effects are estimated. In endogenous samples where treatment effects are "not point identified," the attributes are included as covariates but (point) estimates are not causally identified.

With multiple outcomes, high dimensional treatments, and covariates, the design gives rise to some concerns of limited power, particularly for interaction terms, and of multiple comparisons problems. To alleviate these concerns and make inferences on more theoretically-relevant concepts, I seek to aggregate "up" from the basic AMCE estimates presented here. To test for bias, I make inferences on the basis of F-tests (or the equivalent) on the joint significance of relevant coefficients. To estimate these models, I specify the subset of relevant estimators (β's in Equation 2.12) and implement an F-test to test the null hypothesis that all β's in the subset are equal to zero. I refrain from the use of high-dimensional interactions which are underpowered in the present design. Importantly, note that the inclusion of interactions between identity-based characteristics does not improve the predictive power of the models. Joint tests of interactions between the identity treatments reported in Appendix A.11.3 provide no evidence that identity characteristics serve as complements or substitutes in terms in bureaucrats' responses to petitioners.

2.3.4 On Identification

Section 2.4 shows that bias in effort, Δ, is identifiable by manipulation (random assignment) of petitioner type. The model further implies that there is no direct test of the sources of bias through the manipulated attributes in the factorial design. As such, I test the broader theory about the sources of bureaucratic bias through examination of the testable implications of the model. With one exception, the tests of these implications rely upon treatment-by-covariate interactions with institutional and societal features that I cannot manipulate. While I use a flexible, non-parametric, and interactive covariate adjustment strategy to probe the robustness of these inferences, tests of the comparative statics that use administrative data remain observational.[23]

2.4 Identifying Bias in Bureaucratic Effort

I begin by estimating the magnitude of bureaucratic bias by socioeconomic class, migrant status, and regional accent. Bias in effort, Δ, is identified by *differences* across petitioner treatment conditions. Where this difference is zero, there is no evidence of bias. In addition, baseline *levels* of service provision are also relevant for interpreting these differences.

This analysis focuses on how observed effort varies with randomly assigned petitioner identity characteristics. Given that the regional accent and socioeconomic class (ideally) were revealed as soon as the call was answered, I consider outcomes of process and access as well as the information provision outcomes. I analyze these outcomes on the full sample of answered calls ($n = 1,194$). Logically, the "unrevealed" factors (class, accent, migrant status, and petition difficulty) should be orthogonal to whether or not a call was answered or not. Reassuringly, F-tests of the joint significance of these factors provide no evidence of selection (imbalance) across unrevealed factors on the probability that a call was answered in Appendix A.15.

2.4.1 Bias in Access to the *Alcaldía*

First, I investigate whether petitioner characteristics influence process of the petition through the *alcaldías* in Table 2.5. Column 1 examines whether the dispatcher (original official) identified themselves; if they did not identify themselves ex-ante, callers asked for a name. Levels of iden-

[23] Note that the conditional AMCEs by subgroup are identified. The difference in conditional AMCES – the tests implied by the comparative statics from the model – are not causally identified without imposing additional assumptions.

tification were high (≈ 0.85) and there are no apparent differences across callers per the randomly assigned petitioner characteristics.

Column 2 examines whether the caller was able to ask the question. In general, confederates were unable to make petitions when the dispatcher passed the call or referred the caller to a second official and the second official did not answer within two attempts. The lack of (robust) differences across across identity characteristics is therefore not surprising. Further, it provides no evidence that the dispatcher's handling of calls varied by class or regional accent of the callers.

Column 3 measures whether the petitioner was successful in speaking to (at least) a second official, a measure of access. The mean of this outcome is lower than with the petitions because some petitions were made to the dispatcher directly at the dispatcher's request. There is no bias on the basis of class or accent on this measure of access.

Column 4 examines whether any official identified herself as one of the officials on the pre-treatment administrative lists of MFA and SISBÉN officials collected from government partners. The results indicate somewhat higher levels of access to these program administrators for the middle-class petitioners relative to lower class petitioners, a difference of approximately 4.9 percentage points. Taken with Column 3, this finding likely emerges from higher levels of identification (by a second official) to middle-class petitioners. The joint test of coefficients on class and accent, however, is only marginally significant. Collectively, these analyses suggest limited, if any, bias in navigating the *alcaldías* within an initial interaction on account of class or accent.

The lack of evidence of bias by dispatchers is important for several reasons. First, high rates of identification, petition making, and access to a second official indicate statistical power to identify even modest amounts of bias. The power of the tests combined with the null findings suggest that for this class of tasks, there is no evidence of bias within the present research design. Second, the lack of differences in Columns 1-3 (outcomes measuring dispatcher behavior) provide no evidence that bureaucrats were "differentially confused" by some petitioner or script characteristics.[24]

2.4.2 Bias in Information Provided

Columns 5-10 examine bias in the responses to the petitions. Note that these responses are not conditional on making a petition; thus failing to receive information comprises both wrong responses and no response. Column 5 provides no evidence of bias in the probability that a petitioner receives a complete, correct response on the basis of the identity attributes. Note, however that baseline levels of correct responses are quite low. To the extent that bias represents the *withholding* of effort or information, there is limited scope to move this outcome. In this context, note that the (small) treatment effects on on lower-middle class and resident represent effect sizes of around 20% of this baseline.

There is notable bias in the likelihood of receiving a partial response or any information (Column 6). Lower middle-class petitioners are substantially more likely to receive a partial response or any information relative to lower class petitioners. In Column 7, the point estimate on receipt of any information is 8.1 percentage points and represents a 16 percent increase in the probability of receiving any information relative to the baseline (lower class). There is noisy evidence of a penalty against migrants.

Columns 8 and 9 track two outcomes in which information was not provided. Column 8, "no information" includes any response that did not provide individuals information or invite them to come to the *alcaldía*.[25] These responses included hang-ups, "don't know"-type responses, and situations in which the bureaucrat stated that they did not want to provide information. It is a relatively rare outcome and disproportionately impacts lower-income callers, though the point estimates and

[24]One concern is that because the lower class petitioner scripts were less direct, they may have confused the bureaucrats that answered the phones. There is no evidence that this was the case.

[25]This outcome was not prespecified.

	Access/Process				Response to Petition					
	Dispatcher Gave Name	Petition Made	Second Official	Program Officer	Complete	Incomplete	Any Info.	No Info.	*Alcaldía* Only	Red Tape
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A: IPW ESTIMATES										
Lower-Middle Class	0.011	0.014	−0.014	0.043	0.021	0.048*	0.069**	−0.019	−0.036**	0.003
	(0.020)	(0.025)	(0.028)	(0.026)	(0.018)	(0.029)	(0.029)	(0.021)	(0.017)	(0.024)
Bogotá Accent	0.029	0.038	0.030	0.027	0.002	−0.005	−0.003	0.026	0.015	−0.056*
	(0.026)	(0.031)	(0.034)	(0.033)	(0.022)	(0.035)	(0.035)	(0.025)	(0.019)	(0.030)
Costeño Accent	0.034	0.078**	−0.020	−0.038	−0.007	0.033	0.026	0.020	0.032	−0.027
	(0.026)	(0.031)	(0.035)	(0.032)	(0.022)	(0.036)	(0.036)	(0.026)	(0.021)	(0.030)
Resident	0.015	−0.018	0.015	−0.002	0.027	0.003	0.029	0.007	−0.054***	0.036
	(0.020)	(0.025)	(0.028)	(0.026)	(0.018)	(0.029)	(0.029)	(0.021)	(0.017)	(0.024)
F-test, p-value	0.472	0.071*	0.445	0.078*	0.462	0.352	0.100*	0.698	0.001***	0.197
Placebo F-test, p-value	0.400	0.761	0.643	0.891						
PANEL B: ESTIMATES WITH ENTITY FIXED EFFECTS										
Lower-Middle Class	−0.015	0.016	−0.021	0.046*	0.024	0.057**	0.081***	−0.027	−0.039**	0.002
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
Bogotá Accent	0.023	0.002	−0.023	−0.016	−0.006	−0.044	−0.050	0.039	0.014	−0.098***
	(0.026)	(0.034)	(0.035)	(0.034)	(0.026)	(0.038)	(0.038)	(0.028)	(0.022)	(0.036)
Costeño Accent	0.021	0.038	−0.063*	−0.058*	−0.005	−0.016	−0.021	0.025	0.034	−0.071**
	(0.026)	(0.034)	(0.034)	(0.031)	(0.026)	(0.038)	(0.038)	(0.029)	(0.023)	(0.035)
Resident	0.021	−0.020	0.022	−0.010	0.020	0.021	0.041	0.00005	−0.061***	0.041
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.022)	(0.017)	(0.026)
F-test, p-value	0.612	0.495	0.193	0.060*	0.597	0.153	0.009***	0.447	0.000***	0.015**
Placebo F-test, p-value	0.173	0.693	0.481	0.840						
PANEL C: ESTIMATES WITH ENTITY + ENUMERATOR FIXED EFFECTS										
Lower-Middle Class	−0.016	0.019	−0.026	0.045*	0.028	0.053*	0.081***	−0.023	−0.039**	0.004
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
Resident	0.020	−0.017	0.017	−0.014	0.022	0.019	0.041	0.003	−0.061***	0.040
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
F-test, p-value	0.468	0.263	0.020**	0.061*	0.033**	0.306	0.035**	0.130	0.001***	0.567
Placebo F-test, p-value	0.411	0.246	0.019**	0.166						
Mean DV, Lower Class	0.859	0.758	0.685	0.276	0.104	0.396	0.5	0.153	0.104	0.234
Mean DV, Paisa Accent	0.842	0.723	0.675	0.296	0.119	0.414	0.533	0.124	0.066	0.266
Mean DV, Resident	0.856	0.771	0.668	0.302	0.101	0.428	0.529	0.134	0.108	0.219
Program	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
All Factors	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DV range	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}
Observations	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.5: Bias is estimated by the AMCEs of identity-based characteristics on access to bureaucrats and responses to petitions. All specifications are estimated in OLS with heteroskedasticity-robust standard errors. Panel C omits the accent manipulation because regional accents were voiced by natives of each region (not acted); joint tests in this panel refer only to class and migrant manipulations. The F-tests test the joint significance of coefficients on class and accent factors in Columns 1-4 (prior to the petition) and class, accent, and migrant factors in Columns 5-9. The placebo F-test test the joint significance of attributes of the petition, migrant status and technical question, in Columns 1-4.

F-tests are not significant at conventional thresholds.

Column 9 measures whether or not individuals were simply told to "come to the *alcaldía*" without further information. While all services require the person to come to the *alcaldía* with documents, failure to specify these requirements by phone passes the cost onto the citizen. The estimates suggest that lower-middle class individuals are 37.5 percent less likely to receive this response than lower class individuals, while residents are half as likely as migrants to receive the response.

The results in Column 10 indicate disproportionate use of red tape – a request for *extra* requirements – against Paisas relative to both Bogotanas and Costeñas, with sizable point estimates of 0.071 and 0.098, respectively. These differences are not driven by individual enumerators or pairs of enumerators across the groups (Appendix A.16.2). It is unclear why differences emerge on this outcome specifically.

The observed biases in information provision on the basis of petitioner class merit some additional discussion. It does seem that the class treatment was recognizable; independent coders identified the assigned coding in 77.5% of calls, as reported in Table 2.2.[26] While class is necessarily a compound treatment in the Colombian context, analysis of the magnitude of the "complier" AMCE relative to the intent-to-treat AMCE in Appendix A.16.1 suggests that bias enters through what blinded coders perceive to indicate social class within the calls.[27]

There are several explanations for the generally null effects of regional accent. First, it could be that bureaucrats did not hear regional accents. This seems highly unlikely as blinded coders listening to recordings of the calls correctly identified over 99% of calls (Appendix A.11.1). Second, it could

[26]The blinded coders were given an "I don't know" option in addition to the two class categories; another 13.5% of calls fell into this category. Only 9% of calls were incorrectly classified.

[27]The estimates of the complier AMCE can be seen as an informal test of the excludability assumption as applied to the social class treatment.

be that a Bogotá accent means something different in Bogotá than in other parts of the country. To this end, I report the results of a prespecified analysis on the subsample of calls from the regions in which these accents are local in Appendix A.16.3. Here the treatment is defined as an "in-region" accent. In the full subsample, estimates are near-zero and confidence intervals bound zero for all outcomes. This masks some heterogeneity between the three regions, however. While I cannot reject the null hypothesis that any of the conditional AMCEs is zero, there is suggestive evidence that Costeñas are *punished* in their home region relative to outsiders, while there is mild evidence of an in-region bonus for both Bogotanas and Paisas in their respective home regions.

Collectively the differences in treatment of lower-middle class versus lower class petitioners track those of residents versus migrants, though the class effects are substantively stronger. However, as indicated in Table 2.4, migrant status was not revealed until the petition was made. Appendix A.16.4 reveals that these estimates are conservative and less efficient than estimates of migrant status on the sample of petitions alone. To the extent that these groups are relatively marginalized at least within the experimental comparisons, these comparisons provide some evidence about the dynamics of bias that I explore in the next section.

Beyond the behavioral measures, confederates evaluated their interactions with bureaucrats after each call. These results, reported in Appendix A.16.5, suggest that perceptions largely aligned with the behavioral outcomes. Within enumerator and *alcaldía*, enumerators perceived slightly worse treatment when calling as low-income petitioners. The alignment between the behavioral measurements and perceptions of the calls increases confidence in the behavioral measures.

2.4.3 Does Information Provision Reflect Costly Effort?

I seek to validate that information provision does indeed reflect exertion of costly effort. I consider the total amount of time spent on the call (mean: 4.83, standard deviation: 6.32 minutes). Because the scripts for the petitions varied in length of delivery across the identity characteristics, I lack the ability to identify differences across petitioner identities on this outcome.[28] This represents an excludability violation: observation of a longer call could mean the bureaucrat spent more time answering the question or that the petition took longer to make. By the same token, where script length and information provision counterbalance each other, an inference of no difference in time does not provide a clear measure of differential effort.

Instead, I show that the length of calls is increasing in the amount of information provided (correct, partial, or no information).[29] I first aim to purge differences in the length of calls due to variation in the experimental scripts. To do so, I fit a regression of $\ln$(Minutes on call) on the experimental factors, a program indicator, and enumerator fixed effects with IPW.[30] I then examine the distribution of residuals from this regression across the three types of outcomes.

Figure 2.3 depicts the distribution of residualized (logged) call length by the amount of information provided as empirical CDFs (ECDFs). The graph indicates that the cumulative length of contact for petitions providing no information was substantially shorter than the length of those providing some information. On average, petitions receiving no information were 1.17 minutes shorter ($p < 0.01$) than calls providing partial information and 1.21 minutes ($p < 0.01$) shorter than calls providing complete answers. These differences represent effects of approximately 25 percent of the mean for calls with no information (4.63 minutes). Further, the crossing of the ECDFs for partial and complete information provide some evidence to adjudicate the competence vs. effort distinction between the two types of answers. It suggests that the difference between the two answers is not simply differential competence and that, in the upper median of the distribution, bureaucrats spent more time to provide a more complete answer. This is consistent with qualitative observations of

[28]For example, the migrant petitions included an extra sentence that was not included in the resident petition.

[29]Incorrect includes the *alcaldía* only response from Table 2.5.

[30]There was some heterogeneity in the pacing of calls between enumerators.

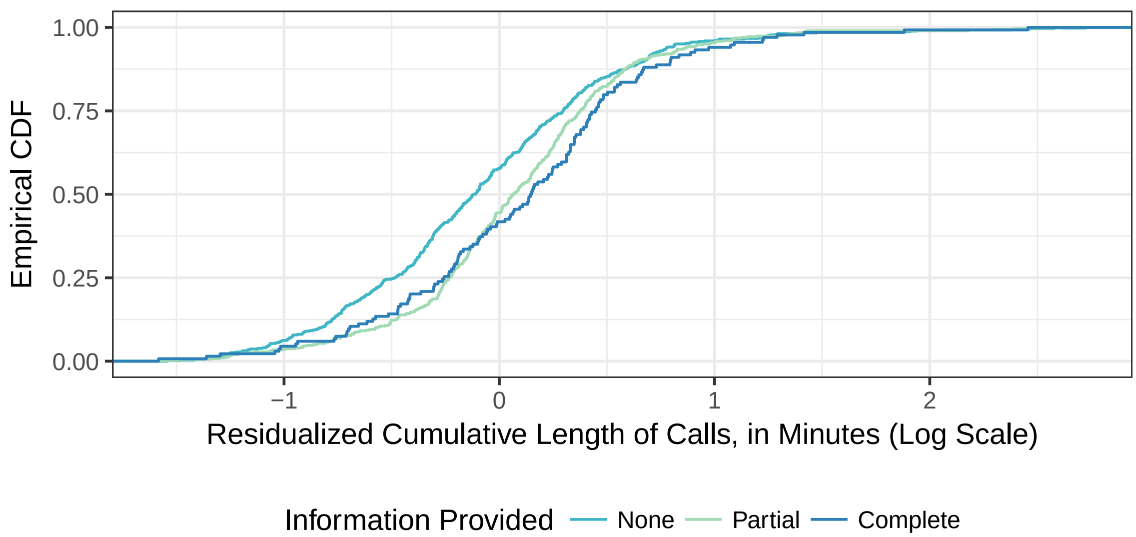


Figure 2.3: The distribution of residualized call lengths by the amount of information contained in the response. Calls that provided no information were uniformly shorter than calls with partial or complete responses. In the upper quantiles of the distribution, calls providing correct answers were longer than those providing partial information.

confederates.

2.5 Examining the Mechanism

The evidence of bias in information provision against lower class petitioners and, to a lesser extent, internal migrants motivates analysis of what drives these biases. I seek to distinguish between the three mechanisms suggested by the model: bureaucrats' tastes, politicians' tastes, and complaint-driven bias. The experiment measures bureaucrats' effort, meaning I do not measure oversight or biases in oversight directly. Indeed, the question is not whether biased oversight leads to biased service outcomes, but how the likelihood of oversight conditions bureaucrats' initial behavior. I use the model to identify the conditions under which oversight-driven bias should be magnified.

To conduct this analysis, I proceed in two steps, following Proposition 3. To begin, I endeavor to separate bureaucrats' taste-based bias from oversight-based bias (composed of politicians' taste-based and complaint-driven bias) on class. Then, I seek to tease apart complaint-driven bias and the politician's taste-based bias. These tests follow directly from the comparative statics presented in Table 2.1. I then distinguish the results from an alternate account from existing models of differential treatment in the form of screening clients for services.

2.5.1 Political Oversight Drives Bias

In order to disentangle bureaucratic taste-based bias from oversight-driven, bias, I consider two parameters of the model: the politician's marginal cost of effort, c_P, and bureaucratic incentives, r. Do the costs of a task reduce the level of bias in effort? As effort becomes more costly to the bureaucrat and audits become more costly for the politician, effort should decline. Increases in these costs should also attenuate bias.

Analysis of bias as a function of petition difficulty provides an experimental test of this logic. I find that class-based bias emerges on easy questions but not on technical questions, where the cost of effort is highest for the bureaucrat. The left panel of Figure 2.4 visualizes estimates from an interactive specification where the technical petition indicator is interacted across all other experimental factors (petitioner characteristics and program) in the design. Consistent with the theory, substantially less information is provided in response to the technical petition. Further, there is no evidence of bias, given that the solid and dotted blue lines are do not substantially diverge for any response. Bias is driven by the easy (registration) petition, as is evident from the divergence of the green lines. Bias is most strongly apparent in easy petitions for the provision of *any* information ($p < 0.002$). Indeed, the difference in the estimated bias against poor petitioners for easy and technical questions in the provision of any information is substantively large at 10.1 percentage points and statistically significant at the $\alpha = .1$ level in a two-tailed test ($p = 0.079$).

This finding is consistent with the prediction with any of the bias mechanisms. However, consider that technical petition may also induce a shock to the politician's cost of effort, c_P. Per Proposition 3, if bias varies in the politician's cost of effort, there is evidence of oversight-driven

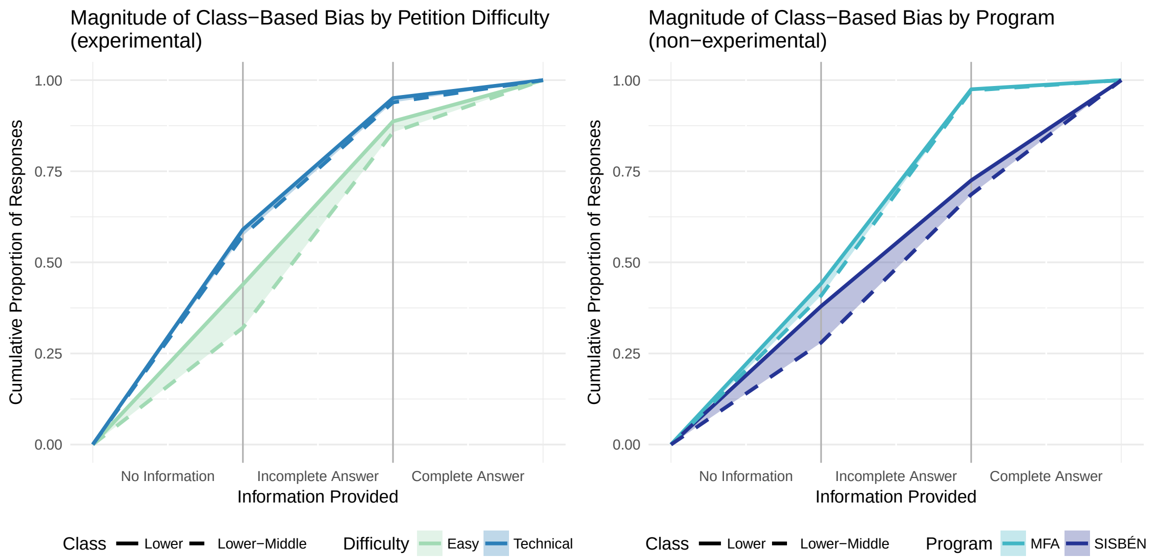


Figure 2.4: Sensitivity of bias in information provision to the cost of effort (difficulty of the petition) (left) and program (SISBÉN or MFA) (right). The area of the polygon corresponds to the level of bias for each subgroup and for each level of information provision. Given that the lower-middle class consistently received more information, the lower curve of all polygons represents the lower-middle class and the upper curve represents the lower middle class. All estimates come from interactive models estimated with IPW.

bias ($\frac{\partial \Delta}{\partial c_P} \neq 0$ if and only if $\Delta_O \neq 0$). If the task is harder for the bureaucrat, it should also be harder for the less expert politician. This (simultaneous) increase in the cost of the politician's effort should increase the relative contribution of the bureaucrat's taste-based bias to the estimated bias in effort. Stated another way, if the oversight-driven biases were entirely absent, the reduction of bias observed in the data corresponds to a very large shock to c_B. Importantly, service provision is not commensurately driven to zero. There is scope to observe taste-based bias, but I do not detect any. This test provides no evidence against oversight-driven bias.

As a further test of whether bias varies in costs to the politician (c_P), consider variation between the two audited programs as a more direct test of the oversight channel. By all accounts, one fundamental distinction between the operation of SISBÉN and MFA at the municipal level is the degree of politicization. One plausible operationalization of politicization is a lower c_P for SISBÉN and a higher c_P for MFA since it is less costly for a politician to intervene in SISBÉN. Examining variation in information provision by program thus provides one test of how bias changes when the politician is more apt to intervene.

For sufficient increases in the politician's cost of effort, she is less willing to monitor the bureaucrat. This implies a higher propensity to monitor the more politicized program (SISBÉN) than the less politicized program (MFA). Less monitoring reduces the bureaucrat's incentives to work and reduces the contribution of oversight-driven bias to total bias. Thus, if bias attenuates substantially for MFA relative to SISBÉN, there is evidence that bias enters through oversight. The results in the right panel of Figure 2.4 mirror this expectation quite precisely. More information was provided for SISBÉN than MFA. There is clear class-based bias in the the provision of any information for SISBÉN ($p = 0.013$) but no evidence of bias in the administration of MFA ($p = 0.421$). The difference-in-difference estimate on the interaction between class and program is sizable at 6.7 percentage points, but is not statistically significant at conventional thresholds ($p = 0.24$). Regression tables supporting these analyses are reported in Appendix A.17.1.

A final analysis seeking to disentangle bureaucratic taste-based from oversight-driven bias turns to variation in bureaucrats to understand bias. Per Proposition 3, an increase in the "bite" of punishment, r, should increase oversight-driven bias. The central characteristic of bureaucratic employment that I measure is contract type, comparing contractors to civil servants. I argue that contractors, all else equal, face higher powered incentives to exert effort given the prospect of contract non-renewal.[31] In terms of the model, I argue that r is higher for contractors than for civil servants. As such, we would expect higher levels of effort and a magnification of any oversight-induced bias.

In this analysis, I study the program officers administering the program in each municipality.

[31] Interviews suggest contractors – patronage or not – work hard, sometimes to "compensate" for shirking civil servant colleagues.

Because Table 2.5 indicated that identifiability of program officers was related to the class of the petitioner, I must restrict analysis to the dispatcher sample or the program officers (from administrative data) to avoid conditioning on a post-treatment variable. Program offices were likelier to take the petition than dispatchers so I focus on the *ex-ante* list of program officers. Table 2.6 examines the conditional effects by program officer contract. It provides no evidence of differences, on average, in bias between contractors and civil servants. Specifically, there are few differences in information provision in the aggregate.[32] Moreover, rates of bias against lower class individuals cannot be distinguished between contractors and civil servants.

The null findings of this analysis are mixed with respect to the model. However, two caveats in the present analysis are in order. As documented in Appendix A.12, I am unable to identify individuals in the poorest municipalities. As subsequent analyses demonstrate, these are precisely the locations where class-based bias is most pronounced. Further, the majority of calls did not reach the individual identified as a program officer in the data. In larger municipalities, they often spoke to subordinates; in smaller municipalities, they spoke to other individuals from the *alcaldía* with or without knowledge of the program. In that sense, this result is very much an "intent-to-treat" effect and attenuate differences between contractors and civil servants to zero.

Second, and more importantly, politicians' choice of hiring mechanism is, to some extent, strategic. While I lack data on the tenure of all SISBÉN program officers, approximately half of all MFA officers (whether contractors or civil servants) have been appointed in the last two years, and fewer than 19% have served more than five years. This suggests substantial scope for mayors' appointment of individuals to the position, whether or not they had previously served in the *alcaldía*. To this extent, the distribution of contract types varies across the two programs. Among the identified

[32]There is evidence that contractors provide more information in response to technical petitions. The estimates in the table refer to an easy petition given the interactions in the estimator.

	Complete (1)	Incomplete (2)	Any Info. (3)	Alcaldía Only (4)	Red Tape (5)
PANEL A: CONDITIONAL AMCE BY EMPLOYEE TYPE; IPW ESTIMATES					
Lower-Middle Class	−0.002	0.077*	0.076*	−0.044*	−0.032
	(0.026)	(0.042)	(0.041)	(0.025)	(0.035)
Contractor: Lower-Middle Class	0.043	−0.040	0.003	−0.003	0.062
	(0.042)	(0.067)	(0.067)	(0.039)	(0.057)
Conditional Effect, Contractor	0.042	0.037	0.079	-0.047	0.030
	(0.034)	(0.053)	(0.054)	(0.030)	(0.045)
PANEL B: CONDITIONAL AMCE BY EMPLOYEE TYPE; STRATUM + ENUMERATOR FE					
Lower-Middle Class	−0.002	0.084**	0.082**	−0.045*	−0.027
	(0.026)	(0.041)	(0.040)	(0.024)	(0.035)
Contractor: Lower-Middle Class	0.042	−0.054	−0.012	0.0002	0.053
	(0.041)	(0.067)	(0.066)	(0.037)	(0.057)
Conditional Effect, Contractor	0.040	0.030	0.070	-0.045	0.026
	(0.032)	(0.053)	(0.053)	(0.029)	(0.044)
Observations	1,194	1,194	1,194	1,194	1,194
Mean, Lower Class and Contractor	0.077	0.438	0.515	0.112	0.207
Mean, Lower Class and Civil Servant	0.125	0.407	0.532	0.115	0.276

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.6: Bias and type of program officer type (civil servant or contractor). Conditional AMCEs of a lower-middle class petitioner by program officer type. The base category is civil servant. In the sample of petitions, there are $n = 611$ petitions to civil servants; $n = 354$ petitions to contractors; $n = 103$ petitions to a vacant program officer; and $n = 126$ petitions to an official whose contract type is unidentifiable. The program indicator is interacted with all factors in the experimental design and a program indicator. Heteroskedasticity-robust standard errors in parentheses.

MFA officials 50% are contractors; just 25% of SISBÉN officials are contractors.[33] Because the effort gains of contractors relative to civil servants are driven by the ratio of r to c_P, the imbalance in contractors should theoretically attenuate differences in the strength of oversight-driven bias in the full sample.

Collectively, these analyses provide evidence that the observed bias is driven, at least in part, by oversight of political principal, e.g. $\Delta_O \neq 0$. The degree to which bias is attenuated by technical questions or less politicization, there is evidence that taste-driven bias by bureaucrats is quite

[33] The fact that there were *any* civil servants running these programs in the municipalities surprised officials that oversee the program at the national level.

limited, e.g. $\Delta_B \approx 0$. This conclusion is supported by several other facts beyond the scope of the model. First, while confederates perceived worse service on average after lower-class calls, they perceived precisely no difference in respect (see Appendix A.16.5). Respect is the only measure of affect in the battery, an indicator of bureaucratic bias in existing audit studies (Einstein and Glick, 2016). Second, the lack of bias in the access outcomes in Table 2.5 provides additional evidence that outcomes that are less easily audited (since complaints are typically received and ratified through the dispatcher) do not exhibit class-based bias.

2.5.2 Bias Occurs where Differences in Access to Complaints is Greatest

In the model, bias between groups is driven by the differentiation of citizens. The differentiation of citizens could come from three sources: differences in the distribution of costs of complaint across groups (η_Q), differences in tastes of the bureaucrat (η_B), differences in tastes of the politician (η_P).[34] Bias should be greatest in magnitude in places where these distances are larger. To examine whether observed biases vary with the distribution of citizen costs of complaint, I leverage the fact that markets for social services vary substantially across Colombia.

I argue that the differences in the distribution of costs of complaint are relative. Measures of physical distance and familiarity with the bureaucracy vary with context. In particular, the relative status of the individual analogues to the experimental profiles – lower and lower-middle class petitioners – varies substantially across Colombia. The intuition is straightforward: a lower-middle class profile connotes a higher status, with more access to complain, in a place where the entire population is poor than a place with many lower-middle class (and higher) individuals.

For concreteness, consider the role of the lower and lower-middle class petitioner profiles in two hypothetical municipalities. In one municipality, the vast majority of the population is poor (lower class). In another, a plurality of the population is lower-middle class. The experimental

[34]To the extent that the previous section focused on separating Δ_B from Δ_O, it implicitly addresses η_B.

profiles relate quite differently to the underlying population distribution. If the cost of complaint is a function of relative *status*, the status differential between experimental profiles is far larger in the first (poorer) municipality than in the second.

Thus, to understand where class-based bias emerges, I examine how class differences in treatment vary with the class composition of municipalities. I assume that differences in cost of access to complain are greatest where the lower middle-class is most empowered relative to poor individuals. This occurs in places with more poor citizens, or higher poverty rates. Thus I use poverty rates to operationalize η_Q. Per Proposition 3, the magnitude of bias should increase in municipal poverty if complaint-driven bias is operative. Accordingly, this analysis should be interpreted as heterogeneous treatment effects with municipal poverty rate as the moderator.

Figure 2.5 examines bias in information provision as a function of a the portion of residents in poverty as per the multidimensional index of poverty, calculated from the 2005 census.[35] The figure shows that anti-poor bias emerges against poor petitioners only in poorer places. The bias is restricted to the inscriptions question (left column) and reception of partial information or the *alcaldía* only response, as described above.

To subject these graphical intuitions to a more rigorous test, I run a series of regression analyses in Appendix A.17.2. I bin the poverty index into terciles to reduce functional form assumptions on the moderator. Because poverty and population are strongly negatively correlated ($\rho = -.61$ in the sample), I include an interactive binned population control with deciles of the estimated 2018 population in a second estimator. Both the moderator, municipal poverty, and the (demeaned) population decile bin controls are interacted across the whole design (all factors and the program indicator).

[35]This index is compiled by the Departamento Administrativo Nacional de Estadísticas (DANE) at the level of rural and urban populations within each municipality. I take the weighted average where weights correspond to the share of urban and rural residents in the population.

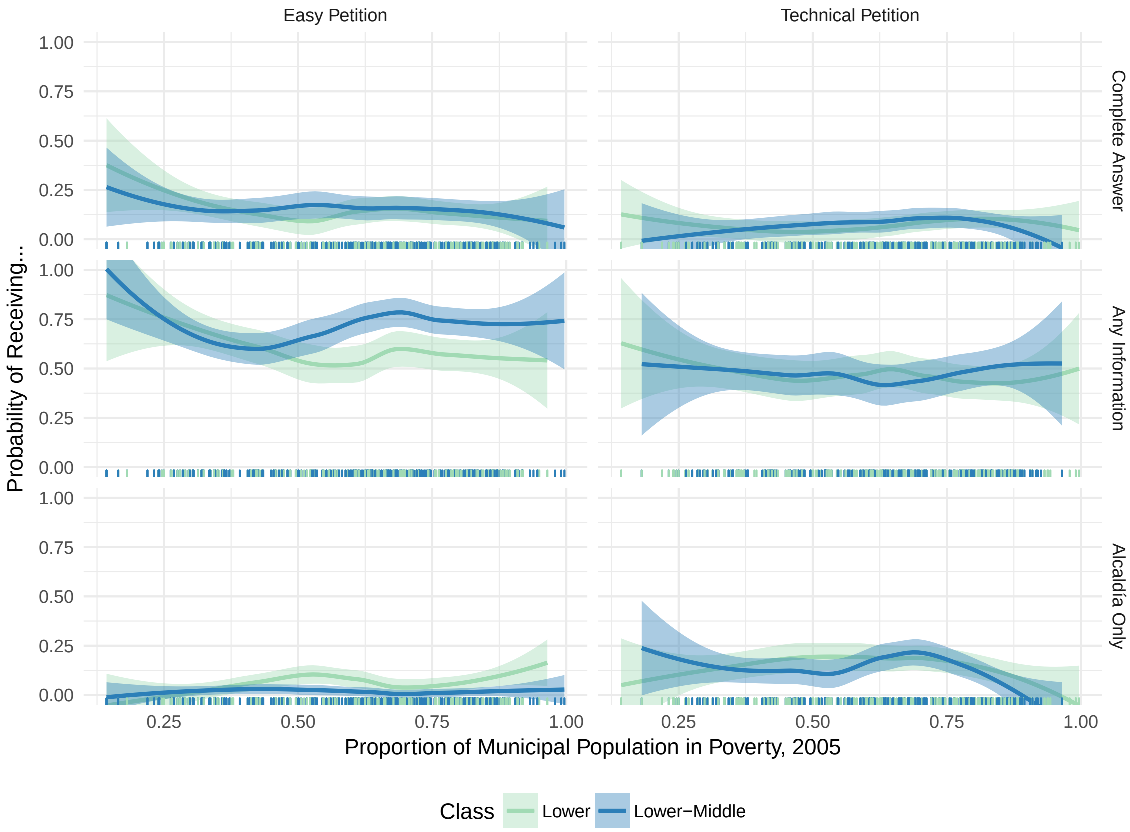


Figure 2.5: Heterogeneity in level of class-based bias by the level of municipal poverty. Column 1 examines average marginal effects on "easy" (enrollment) questions while Column 2 examines average marginal effects on technical questions. Lines are estimated by Loess regression with a span of 0.75. The shaded regions are 95% confidence intervals. Graphs exclude observations from Bogotá given the vastly disproportionate number of observations, though substantive results on bias are robust to including all observations in the analysis.

This analysis suggests that bias against lower-class individuals (the baseline) is worse in poorer places. There is no evidence of bias in the lowest tercile (the municipalities with the lowest poverty rates) for any outcome. Class bias against the poor is increasing in the middle-poverty and high-poverty terciles. I find clear, statistically significant evidence of bias in the high-poverty tercile for the receipt of partial information. Further there is suggestive evidence that differential application of the "*alcaldía* only" outcome against poor individuals is driven by poorer municipalities. These findings are robust to other operationalizations of poverty including rates of secondary education

(2005).

Combined with the analysis of in-region accents, these findings provide suggestive evidence that differentiation of citizens correlates with bias in service provision. Attributing such patterns to tastes requires a theory of bureaucratic selection or politician incentives that yield divergent tastes. While I cannot eliminate this possibility, several findings are useful to consider. First, there is no evidence that service favors the median voter in each municipality – neither the class or regional accent analysis supports this interpretation. If service were to favor the median voter, the poor should do the *best* in the highest poverty places; these are the places that they do the *worst*. From an elected politician's perspective, providing worse service to poorer individuals in poor places works against the median voter and those most likely to turn out in Colombia (Kasara and Suryanarayan, 2015).

Other explanations of bias in terms of politician tastes do not account for these geographic patterns of bias. While there may be a disproportionate incentive to politicize social programs (outside the scope of the model) to claim credit or buy votes in poorer places, it is not clear why such opportunities to claim credit would yield *unequal* information provision, as opposed to simply less information provision. Following accounts of clientelistic usurpation of social services in Weitz-Shapiro (2012), we would expect clientelism to correlate with lower *levels* of service. Importantly, as is evident in Figure 2.5, there is no evidence that less information is provided to the middle class in poor municipalities. Further, as I document in A17.2, clientelism practices are highly regional in Colombia; this pattern persists *within* region and department. Finally, I leverage the municipal classification of electoral risk including clientelism, corruption, and electoral violence by Colombia's Mission of Electoral Observers (MOE) to show that these patterns persist when interactively controlling for these features (Misión de Observación Electoral, 2018).

Unless politician tastes vary systematically in unmeasured ways with the degree of poverty in a municipality, there is little evidence supportive of politicians' taste-based bias in driving the bias results. I find no evidence that political competition drives bias. If competition drives politician incentives to provide public goods, S is the relevant parameter of the model. Note that the expression for bias in effort (Equation 2.8) does not include S, implying S does not drive bias. However, relaxing the assumption of interior effort, bias is eliminated when S drives universal service provision. Empirically, the low rates of information provision suggest that this is not the case for the relevant services. To the extent that competition drives the selection of different "types" of politicians, there is no evidence that this manifests in politicians with different tastes.

One final explanation concerns the selection of bureaucrats themselves. In rural areas, in particular, bureaucrats are hired from a less skilled local labor market. This could drive anti-poor bias via lower competence, perhaps accompanied by tighter oversight, or different tastes, potentially as a function of status. To this extent, a theory of bureaucratic selection consistent with these results cannot distinguish between oversight and taste explanations. Further, recall that the estimates in Appendix A.17.2 suggest that the association between poverty and bias is not driven by differences in population.

I therefore argue that the most plausible interpretation of the finding is that where differentials in relative ability to complain between the treatment conditions are theoretically the strongest, levels of bias against the less able group are strongest. Structurally, this analysis suggests that $\eta_Q > 0$. In particular, lower-middle class individuals are relatively more empowered in places where a plurality of the population is poor. Suggestive patterns of an in-region penalty against Costeñas in the Coast (Caribe) are consistent with this interpretation. If Costeños are perceived as less demanding and therefore less likely to complain within the region, the observed penalty is consistent with the complaint-driven bias explanation of findings.

2.5.3 Alternative Explanation of Bias: Screening

The theoretical model described here is a model of underprovsion rather than misallocation of services. Some theories of misallocation suggest that the "bias" that we observe could simply be efforts to screen citizens that the program is intended to serve from unintended potential recipients (Banerjee, 1997; Ting, 2017). The intended population of beneficiaries/registrants for SISBÉN and MFA are poor individuals and households. However, I find no evidence that differences in levels of service provision are consistent with the theoretical predictions of a screening account.

Informational outcomes cut *against* lower class individuals and internal migrants, the target populations for these programs.[36] One outcome of interest for testing the screening logic is the use of red tape. Red tape is hypothesized to serve as a mechanism that induces individuals to truthfully reveal their "type" – whether or not they comprise the intended target population of the program. Table 2.5 indicates high levels of red tape generally, but no differences in its application by class or migrant status.

While a story of screening is inconsistent with the observed data, it may still be the case that red tape is employed to deter unintended recipients from requesting services in the first place. Yet, qualitatively, the forms of red tape requested appear to disparately impact intended recipients (the poor). The most common extra requirements were a receipt for utilities (usually electricity), a formal letter of application for the service, or extra government documents (other services). Thus, within the experimental data, there is no evidence of differential treatment due to screening. Speculatively, in a setting with endogenous requests, the type of red tape may exacerbate inequality in service provision, but in a direction *opposite* to that predicted by existing screening theories.

[36]To the extent that migrants requesting services are associated with internally displaced persons (IDPs), a special category for both programs, these programs should also favor the migrant condition.

2.6 Discussion: Bias in Effort and Inequality in Outputs

To what extent does bias in information provision map onto inequality in public service outputs? Bureaucratic bias in effort is important because of its link to inequality in citizen access to public services. Recall that the theory suggests that in the presence of oversight-driven bias, bias in effort is a sufficient condition for ultimate inequality in outcomes.

While the experiment allows for measurement of bureaucratic effort in information provision, confederates did not try to obtain the service. Yet, using pretreatment data on SISBÉN registration from across Colombia's municipalities reported in Figure 2.1, I examine the correspondence between rates of enrollment (outputs) and the experimental measures of bias.[37] Recall that two pathologies of SISBÉN enrollment exist in the administrative data: over enrollment and under enrollment of the relevant population. Some municipalities maintain rolls that could not possibly cover the estimated poor population; other municipalities maintain rolls far larger than the population as a whole. The focus of this experiment is on the former category. To that end, I investigate whether there exists differential levels of bias in municipalities where the service is under-provided from places in which it is plausibly administered according to program guidelines.

Table 2.7 suggests that bias is present precisely in the places in which under enrollment of plausible beneficiaries is the strongest concern.[38] There is strong evidence of bias in information provision in the base category (under-enrolled) municipalities. This bias is substantively, and for some outcomes, significantly attenuated in municipalities with ostensibly "intended" enrollment. These results are robust to redefinition of the "plausible enrollment" category (see Appendix A.17). While it is evident that under-enrollment occurs in poorer places, the results are robust to controlling

[37] MFA data by municipality is not publicly available. However, aside from IDPs and indigenous Colombians, SISBÉN is used to qualify for MFA. As such, under-enrollment of SISBÉN should predict under-enrollment of MFA.

[38] To the extent that over-enrollment represents politicization, empirically bias also emerges in these places.

	Complete	Incomplete	Any Information	*Alcaldía* Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
PANEL A: CONDITIONAL AMCE ON CLASS BIAS BY ENROLLMENT TYPE					
Lower-Middle Class	0.072	0.091	0.163***	−0.039	−0.010
	(0.045)	(0.064)	(0.061)	(0.049)	(0.055)
Plausible Enrollment: Lower-Middle Class	−0.056	−0.072	−0.129*	−0.005	0.006
	(0.050)	(0.074)	(0.071)	(0.054)	(0.064)
Conditional Effect, Plausible Enrollment	0.015	0.019	0.034	-0.044**	-0.005
	(0.022)	(0.037)	(0.036)	(0.022)	(0.033)
PANEL B: CONDITIONAL AMCE ON CLASS BIAS BY ENROLLMENT TYPE WITH COVARIATES					
Lower-Middle Class	0.108	0.082	0.190**	0.008	0.076
	(0.051)	(0.092)	(0.089)	(0.058)	(0.082)
Plausible Enrollment: Lower-Middle Class	−0.094	−0.084	−0.179*	−0.062	−0.112
	(0.059)	(0.108)	(0.105)	(0.069)	(0.097)
Conditional Effect, Plausible Enrollment	0.013	-0.002	0.011	-0.053**	-0.036
	(0.027)	(0.042)	(0.040)	(0.024)	(0.036)
Interactive Poverty Decile Bins	✓	✓	✓	✓	✓
Interactive Poverty Decile Bins	✓	✓	✓	✓	✓
Mean, Lower Class and Plausible Enrollment	0.107	0.449	0.556	0.107	0.251
Mean, Lower Class and Under Enrollment	0.084	0.379	0.463	0.126	0.242
Observations	903	903	903	903	903
All Factors	✓	✓	✓	✓	✓
Program	✓	✓	✓	✓	✓

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.7: Relationship between bias in information provision and underprovision of SISBÉN. OLS estimates of the conditional AMCE of class by municipal SISBÉN enrollment type. The sample includes places that are under enrolled or plausibly enrolled as intended. Standard errors are clustered by municipality ($n = 366$) because the conditioning variable is measured at the municipal level.

flexibly interactively for municipal poverty and population (Panel B). This finding is consistent with the logic that bias in bureaucratic effort yields inequality in service provision. It bolsters confidence that the bias in effort measured in the experiment correlates with public service outputs. These results are also consistent with the theoretical extension of endogenous requests for service. In places with where prospects for service are lowest, lower-income Colombians may opt out of seeking SISBÉN registration altogether.

2.7 Conclusion

Observers of Latin American social policy regularly identify variation in the implementation of programs intended to reduce inequality. While existing literature has emphasized electoral motives of politicians in program implementation or non-implementation (Holland, 2015; Niedzwiecki, 2018), I argue that disparities in administration emerge in the course of everyday processes of service provision even without such political directives. By characterizing service provision as a strategic relationship between a politician, a bureaucrat, and a citizen, I identify a new mechanism through which political inequality in ability to draw oversight from a politician leads to inequality in access to social programs. Empirically, I show that bias against lower-class petitioners in the provision of information is substantial but occurs only where oversight is most likely and in municipalities where inequalities in voice are apt to be strongest.

The model of service provision posited in this paper is broadly applicable beyond the reaches of social services in Latin America. The three sources of bureaucratic bias that I identify should emerge differentially across contexts. The model helps to guide our understanding of the institutional and social conditions under which we observe the complaint-driven biases present in Colombia. In particular, as bureaucrats become more insulated, such bias attenuates. Further, where disparities in voice are lower, the scope for complaint-driven bias decreases.

One implication of this argument is that inequalities in political voice reduce the efficacy of the state programs to combat inequality. I posit that mundane processes of service provision may contribute to inequality traps in highly unequal societies. This mechanism complements literature linking economic inequality to political inequality through more explicit conflict between the interests of elites and non-elites (Acemoglu et al., 2008; Acemoglu and Robinson, 2008).

The argument presented here suggests some policy implications for reducing inequality in access to services. One argument revolves around the *selection* of politicians and bureaucrats. My model implies that selection of politicians or bureaucrats whose tastes favor the poor can offset complaint-driven bias while improving service provision. On the bureaucrat side, this counters normative ideals of neutral bureaucrats. In contrast to selection-based remedies, strategies to combat biased oversight by politicians have fewer prospects for success. While insulating bureaucrats reduces complaint-driven inequalities, it also reduces effort leading to lower aggregate levels of service provision. Further, inducing the politician to monitor service provision to populations unable to complain at higher rates will not eliminate inequality in access if citizens must pay a cost to request service in the first place. In general, citizen-focused interventions to reduce costs of engaging the bureaucracy hold the most promise for improving service provision and reducing inequality, despite difficulties in implementation. Simply providing information on government services does not remove cultural, economic, and psychological barriers to demanding equal service.

The implications of the theory proposed in this paper provide ideas prime for further exploration. To the distributive politics literature, it suggests that we must consider the production of public goods – not simply budgets – to understand variation in "who gets what." To the bureaucratic politics literature, I argue that citizens can combat the moral hazard of bureaucrats, but that increased bureaucratic effort on the basis of citizen complaints translates to unequal gains in terms of service provision. Collectively, these points suggest that the strategic relationship identified in this paper between politicians, bureaucrats, and citizens has the potential to illuminate big questions about accountability, inequality, and redistribution.

Chapter 3

Oversight, Capacity, and Inequality

Citizen complaint systems represent one of the most regular forms of citizen engagement with governments in democratic and autocratic regimes alike. These systems, broadly defined, provide a means for citizens to convey information to the government in response to a (perceived) failing of a bureaucrat or bureaucratic agency. Across contexts, citizens communicate information about the location of potholes, missing social benefits, corruption by state agents, and violations of social or human rights, among many others.

Despite their ubiquity, the design of citizen complaint systems as a bureaucratic oversight institution varies across contexts and policy areas. These systems differ in how principals (politicians or higher- level bureaucrats) respond to complaints, both in terms of granting redress an in punishing the offending agents. In turn, the anticipation of these responses is thought to shape citizens' complaint-making behavior and bureaucrats' effort in providing services. In this paper, I ask how the design of oversight institutions influences "who gets what" from the state.

A growing literature uses citizen complaints to study government responsiveness (Chen, Pan, and Xu, 2015; Christensen and Ejdemyr, 2020; Sjoberg, Mellon, and Peixoto, 2017; Dipoppa and Grossman, 2020; Hamel and Holliday, 2019) and the organization of autocratic regimes (Pan and Chen, 2018; Dimitrov, 2013). This paper complements and builds upon this literature by considering when and how politicians (principals) design complaint systems and how these institutions shape the population that complains. In turn, this characterization of selection into complaint-making generates new implications for understanding the distributional consequences of oversight.

Consider a politician's choice to use (or ignore) information generated by citizen complaints as part of a bureaucratic oversight strategy. In terms of classic oversight parlance, when do politicians commit to monitor bureaucrats via "fire alarms" versus "police patrols"? How the does choice of such monitoring propensities constrain the politician's ability to incentivize bureaucratic effort? I contend that the answer to this puzzle lies in *who* complains. The model emphasizes that citizens in a population vary in their ability to complain or "pull" a fire alarm. The observation that costs of complaint are often non-trivial and can vary substantially across a population echoes early warnings of McCubbins and Schwartz (1984).

I answer these questions by developing a model of service provision built upon a framework developed by Prendergast (2003). In the model, a bureaucrat chooses whether or not to exert effort to determine a citizen's state (i.e., eligible or ineligible). The bureaucrat then makes a binary allocation of the service to the citizen. The citizen, who knows her state, observes the allocation, and decides whether or not to complain to the politician. Importantly, the cost of complaint varies across the population of citizens, such that citizens vary in their willingness to provide information about their personal state. The politician monitors the bureaucrat's action based on the allocation and the presence of a complaint. If such auditing reveals misallocation (i.e., a service denial to an eligible citizen), the citizen recovers the service and the bureaucrat is punished.

The politician designs oversight by committing to a contract *ex-ante* that specifies effort incentives for bureaucrats (the magnitude of punishment for errors) and monitoring rates as a function of observed allocation by the bureaucrat and the presence of a citizen complaint. I characterize four qualitatively distinct contracts that emerge in equilibrium depending on: (i) the politician's targeting of services and (ii) the level of bureaucratic insulation, conceived as a limit on the allowable magnitude of effort incentives. The contracts vary in their provision of effort incentives and whether the politician's monitoring responds to information provided by citizens in the form of complaints. These monitoring propensities, in turn, determine which citizens have an incentive to complain.

I use these oversight contracts to derive implications for the state's capacity to implement policies as well as for inequality in access to services. The measure of implementation capacity developed in this paper arguably formalizes Mann's (1984) concept of state capacity as "infrustructural power," or the "ability to . . . penetrate civil society, and to implement political decisions throughout the realm" (189).[1] By focusing on an informational problem underlying service provision, namely the need for a government to learn about its citizens, this paper draw parallels to discussion of the "legibility" of citizens to a government as a determinant of implementation capacity (Scott, 1998; Lee and Zhang, 2016).

I find that contracts that condition monitoring on information provided by citizens have an ambiguous effect on a state's capacity to match policies to intended recipients. This type of monitoring improves the accuracy of targeting among "legible" citizens – those that choose to provide information when wrongly denied. However, it also promotes a form of capture by simultaneously reducing the state's accuracy in providing service to "illegible" citizens who never complain. Which effect dominates depends on the share of legible citizens in a population.

In contrast to its ambiguous effect on state capacity, conditioning monitoring on citizen information transmission always increases inequality in the delivery of state services when citizens vary in their propensity to complain. When monitoring relies on citizen complaints, citizens who complain receive more accurate – and simply more – services, both from bureaucrats' initial allocations and

[1]There are many definitions of state capacity in the literature. I refer to "implementation capacity" to refer to the concept used in this paper.

through redress of their complaints. Further, the capture mechanism implies that those who cannot complain are worse off (in absolute terms) than they are in the absence of information transmission. The magnitude of this increase in inequality depends on the distribution of costs of complaint in the population.

The use of citizen information in bureaucratic oversight thus suggests a possible tradeoff between expanding the state's capacity to accurately serve its citizens and entrenching inequality in access to state resources. The tradeoff emerges when oversight institutions induce a sufficient share of the population to provide information to the state. Characterization of this tradeoff helps to reconcile divergent arguments about the welfare effects of increased state capacity (i.e. Scott, 1998; Acemoglu, García-Jimeno, and Robinson, 2015; Johnson and Koyama, 2017). It also forwards a need to examine the distributional consequences of efforts to expand capacity.

This paper analyzes a comparatively neglected tool used by politicians to influence the state's capacity to implement policies: the use of information from citizens in bureaucratic oversight (Berwick and Christia, 2018). Existing work linking bureaucratic institutions to state outputs has focused on the adoption of civil service reforms (Geddes, 1991, 1994; Grindle, 2012; Huber and Ting, 2020). The present model captures public sector personnel systems with two exogenous parameters: bureaucratic quality and insulation. In so doing, it allows for consideration of how these oft-studied features of public sector personnel systems affect the oversight schemes adopted by politicians and their consequences. Whereas civil service reforms are often viewed as major, costly reforms across large portions of the bureaucracy (Rauch, 1995; Folke, Hirano, and Snyder, 2011; Ujhelyi, 2014), oversight practices can, in principle, be deployed or manipulated more flexibly by politicians. Variation in oversight practices across policies or jurisdictions therefore may provide more explanatory power for variation in apparent implementation capacity across space, time, and policy design, in line with a growing literature on sub-national variation in capacity (Weber, 1976;

Enriquez and Centeno, 2012; Soifer, 2015).

The model in this paper builds upon an emerging theoretical literature on state capacity (Besley and Persson, 2010; Acemoglu, García-Jimeno, and Robinson, 2015; Gennaioli and Voth, 2015; Snowberg and Ting, 2019). As in the empirical literature, there is no apparent consensus on what state capacity means. I focus on one manifestation of state capacity distinct from the aforementioned literature: the congruence between policies and their realization. I refer to this as implementation capacity. Further, this paper clarifies and formalizes the distinction between bureaucratic capacity and state implementation capacity, by emphasizing the incorporation of bureaucrats into the state as an organization consisting of a government *and* citizens.

This primary contribution of this paper is its suggestion of a link between organization of bureaucratic oversight and the study of distributive politics. The study of "who gets what" from the state generally focuses on the allocation and policy decisions made by politicians by examining which individuals or groups are targeted as beneficiaries (Golden and Min, 2013). One interpretation of existing arguments of capacity and economic growth is that capacity scales the "size of the pie" that politicians have to distribute. My theory, instead, contends that building the capacity to implement policies redistributes the pie across different segments of a population. Such distributional consequences of implementation can occur independently from the content of the policy. To the extent that politicians design influence bureaucratic oversight institutions, I identify a novel strategy via which politicians influence "who gets what" beyond the policymaking process, complementing an emerging literature on policy implementation (Williams, 2017).

Many recent and ongoing studies on state capacity are motivated by an impulse to learn "how to strengthen [state capacity]" (Berwick and Christia, 2018: p. 71). The bureaucratic oversight institutions I study provide a new framework through which to answer the "how." However, the main findings on capacity and distribution suggest that viewing capacity as an aggregate concept or measure can disguise stark distributional outcomes of efforts to strengthen states' implementation capacity. In so doing, I provide one possible reconciliation of a much longer-standing disagreement about the welfare effects of state capacity (e.g., Scott, 1998; Johnson and Koyama, 2017).

3.1 Empirical Motivation

3.1.1 Describing Observed Complaints

The most frequently documented citizen complaints come from 311-type hotlines or online platforms that allow for reporting about a variety of service provision issues (Chen, Pan, and Xu, 2015; Pan and Chen, 2018; Sjoberg, Mellon, and Peixoto, 2017; Christensen and Ejdemyr, 2020; Dipoppa and Grossman, 2020; Hamel and Holliday, 2019). In Figure 3.1, I examine per-capita utilization of complaint hotlines in New York City, United States and Bogotá, Colombia. Similar to many other cities and countries with 311-type systems, both cities release anonymized complaint-level records. The left panel of Figure 3.1 suggests a consistent stream of complaints in both cities between January 2017 and June 2018. Rates of complaint are substantially higher in New York, averaging 616 per million residents per day versus 15 in Bogotá.[2] Variation in the services covered by complaint systems, modes of complaint, and potentially responsiveness render the comparison of complaints in both cities a challenging endeavor. Nevertheless, the non-trivial rates of complaint in both cities suggest that responding to and remedying complaints occupies one source of oversight effort.

The right panel of Figure 3.1 examines rates of complaint across smaller geographic units. The data is shared at different levels of spatial aggregation. In Bogotá, I examine the city's 20 localities; in New York, I examine 2164 census tracts. In both cities, there is substantial variation in the per-capita use of complaints across these areas. The variation across geographic units is substantial: in Bogotá, moving from the first to the third quartile locality represents a 232% increase in the

[2] I exclude non-emergency complaints to the police, fire department, and Taxi and Limousine Commission in New York given to increase comparability. Non-emergency police calls are directed elsewhere in Colombia.

per-capita rate of complaint; in New York, the analogous shift represents a 47% increase in the per-capita rate of complaint. Existing studies that seek to identify differential *responses* to complaints as a function of election timing (Dipoppa and Grossman, 2020), politician re-election incentives (Christensen and Ejdemyr, 2020), or neighborhood characteristics (Hamel and Holliday, 2019) find measurable differences in the speed with which complaints are remedied, but effect sizes are quite small, ranging from a few hours to 2 days. Combined, substantial differences in the rate of complaint and small (if precise) differences in response to complaints suggest that selection into complaining may be particularly important for understanding the distributive consequences of these types of complaint systems.

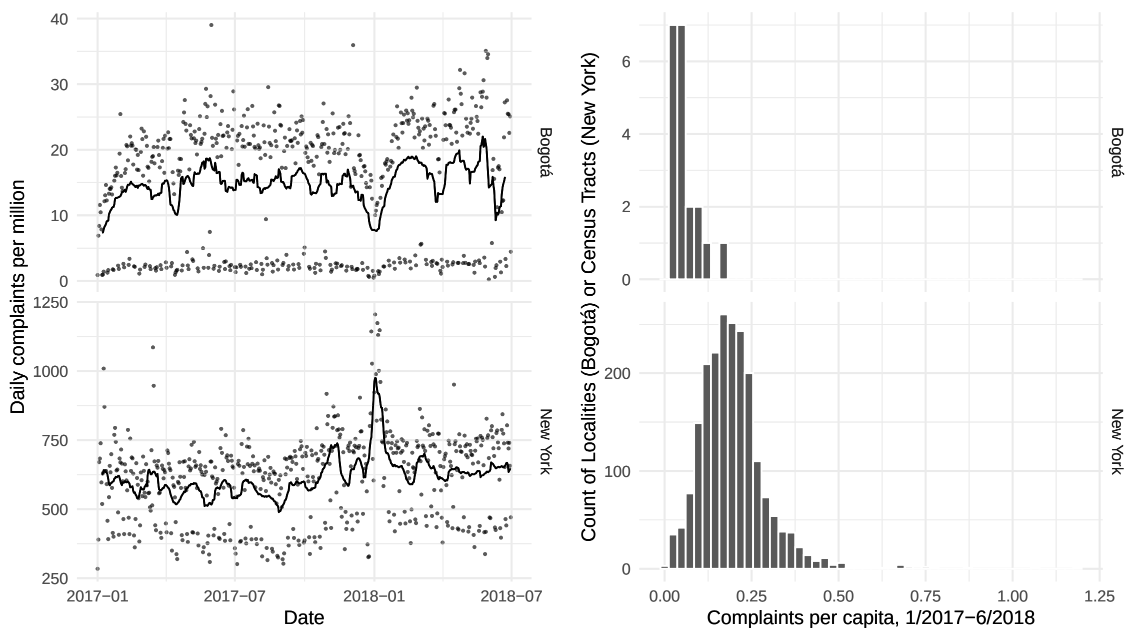


Figure 3.1: Description of anonymized 311 complaint data in Bogotá and New York. The left panel depicts the number of daily complaints per million over an 18-month period. The large differences in daily complaint rates in both cities generally map to weekdays vs. weekends/holidays. As such, the line represents a 7-day moving average. The right panel shows the distribution of per-capita complaints over the 18-month period by locality (Bogotá) and census tract (New York).

Certainly, rates of complaint may differ across a population for multiple reasons. If different subpopulations rely more or less on specific public services, their need for recourse via complaint may vary. With high levels of residential segregation, these differences may manifest across jurisdictions. Further, even among widely-used services, if the quality or level of service provision is uneven across jurisdictions, worse service provision may yield more complaints. On the other hand, if some citizens may face fewer costs or barriers to complaint they may be more likely to communicate grievances that do occur (Ba, 2018; Rizzo, 2019). In the model, as in Chapter 2, I consider a population that is differentially willing (or able) to engage the state via complaint. In light of the complaint data, I provide suggestive evidence in favor of the plausibility of this argument in Appendix B.2.

3.1.2 The Design of Oversight

The design of oversight practices varies substantially across contexts and policy areas. One source of variation evident even across 311 complaints is how costly complaints are to the bureaucrat subject to the complaint. Across national contexts, complaints about a pothole may help to redress the issue but are unlikely to result in a substantial penalty to a bureaucrat with responsibility for roads. However, bureaucrats' efforts to hide corruption complaints in China suggest a widespread perception that these complaints are detrimental to the career advancement of implicated officials (Pan and Chen, 2018). The design of oversight systems includes the determination of sanctions for implicated bureaucrats. Of course, politicians (principals) operate within the constraints of personnel laws in determining bureaucratic sanctions. Where the use of sanctions is permitted, setting these sanctions can be consequential for understanding bureaucratic behavior.

A second important feature of oversight systems concerns how politicians opt to collect and use the information generated by complaints. Relative to other complaint systems, 311 systems impose comparatively low barriers to register a complaint. However, variation in perceived levels of redress appears to condition citizens' decisions to complain (Sjoberg, Mellon, and Peixoto, 2017; Dipoppa and Grossman, 2020). Importantly, oversight systems specify how politicians monitor bureaucrats in the presence and absence of complaints. While I am unaware of systematic classifications of complaint mechanisms and monitoring practices, several anecdotes are instructive.

In Colombia, for example, the 1991 Constitution specifies a process for complaint and appeal of most services beyond the scope of 311 hotlines. Citizens submit written petitions (*peticiones*) for information or recourse to authorities overseeing the relevant entity.[3] Petitions impose costs because they must be formal letters delivered to relevant entities. When responses are not received within the specified timeframe or need to be appealed, complainants can pursue an *acción de tutela*, a legal petition directed to courts for adjudication. These mechanisms are widely, if unevenly, used: 607,499 *tutela* claims were filed with the courts in 2018, or 1.2 cases per 100 citizens (Defensoría del Pueblo de Colombia, 2019; Taylor, 2018). In the public health system where *tutelas* are used most widely, these complaint processes are particularly important drivers of who accesses health services. In this case, the Colombian Constitution guided the adoption of new mechanisms for complaints and specifies protocols for a response.

Other oversight systems develop endogenously without similar "paper" (or constitutional) provisions. For example, Slough and Fariss (2020) examine illegal pretrial detention in the Haitian criminal justice system. They consider a population of unrepresented and illegally detained citizens who lack access to any complaint mechanisms. One implication of the argument is that a lack of access to complaint substantially reduces the rate at which the justice system processes cases relative to represent cases. The oversight institutions in practice vary substantially from the "paper" laws specified by the Haitian Constitution, yet still reflect the choices of a government overseeing the courts. These anecdotes suggest a plausibly broad range of modes via which oversight institutions are adopted and posit some challenges for the empirical measurement of these decisions. Neverthe-

[3]Most entities are public sector agencies, others are public-private entities that oversee service delivery, most commonly in health insurers.

less, they also point to how consequential these decisions can be in terms of access to high-stakes services/outcomes they govern.

3.2 Model

The model examines the choice of bureaucratic oversight institutions in a service provision setting. The model of service provision builds upon Prendergast (2003), with two central departures discussed at length in the model exposition.

I consider three actors: a citizen (C), a bureaucrat (B), and a politician (P). Define a state, $\omega \in \{0, 1\}$, where $\omega = 1$ with probability $\frac{1}{2}$. The state can be thought of as a characteristic of a citizen, specific to a single interaction, as opposed to a fixed characteristic of the citizen (e.g., type). In various service provision settings, the state could refer to sick or healthy; guilty or innocent; or eligible or ineligible. The state is private information to the citizen.

In this model, implementation capacity refers to the congruence between the ultimate service outcome, $a^\dagger \in \{0, 1\}$ and a citizen's state. I denote this congruence C in Equation 3.1. In existing work including Prendergast (2003), congruence is often assumed to generate a social surplus. I abstract from the assumption that accurate targeting of a service (higher congruence) leads to "better" outcomes. Like the motivating examples, accurate targeting of a service could well promote the rule of law or public health. On the other hand, the targeting of state repression is also a manifestation of capacity, but is not typically assumed to generate a social surplus. The framework developed here allows for the comparative study of capacity across society under weaker normative assumptions about the outcomes of state services.

$$C = \begin{cases} 1 & \text{if } \omega = a^\dagger \\ 0 & \text{else} \end{cases} \tag{3.1}$$

The bureaucrat is tasked with providing a citizen with some service, allocating a. He determines whether to exert effort, $e \in \{0, 1\}$, to try to ascertain ω. Exerting effort ($e = 1$) incurs cost $d > 0$. The bureaucrat correctly ascertains the state with probability $q + pe$ where $q \in [\frac{1}{2}, 1]$ and $p \in [0, 1-q]$. The parameter q should be interpreted as a measure of bureaucratic quality and $q+pe$ can be interpreted as the measure of bureaucratic capacity that incorporates both quality and effort.

Upon observation of their allocation, the citizen determines whether to complain ($c = 1$) or not ($c = 0$) to the politician about their allocation, at cost θ, where θ is a random variable distributed according to the pdf $f(\cdot)$, with cdf $F(\cdot)$, where $F(0) = 0$. θ represents the legal, expertise, time, and/or psychological costs of contesting the bureaucrat's allocation. θ can be thought of as an individual citizen's type and is common knowledge. The invocation of a continuum of costs of complaints represents the first substantial departure from Prendergast (2003). Prendergast (2003) assumes that complaints are costless. A direct implication of this assumption is that, conditional on the state and allocation received, there is no variation in citizens' equilibrium complaint-making behavior. This implication stands at odds with recent work measuring citizen participation in service provision (Bussell, 2019; Kruks-Wisner, 2018). Further, consistent with descriptions of legibility before the state, the continuum of costs of complaint suggests that citizens are not equal in their ability (or propensity) to generate information. Importantly, the properties of the distribution of θ serve as a way to consider societies in comparative perspective.

The politician observes the bureaucrat's allocation and the citizen's complaint (resp. non-complaint) and audits the bureaucrat's decision according to a pre-specified contract that stipulates the rate of auditing as a function of a and c. Denote this rate $\rho(a, c) \in [0, 1]$. If audited, the politician pays a cost, $\frac{\rho(a,c)^2}{2}$, to learn ω. If the politician audits and observes that $a = \omega$, she will not change the allocation. If $\omega \neq a$, the citizen's ultimate allocation is $1 - a$. Thus, the ultimate allocation of the service, $a^\dagger$ is given by:

$$a^{\dagger} = \begin{cases} 1 - a & \text{if politician monitors and } \omega \neq a \\ a & \text{else} \end{cases} \tag{3.2}$$

When the politician reverses a bureaucrat's allocation, the bureaucrat is sanctioned with a penalty of $\Delta \in [0, \overline{\Delta}]$. $\overline{\Delta} \in \{\overline{\Delta}_L, \overline{\Delta}_M, \overline{\Delta}_H\}$ is an exogenous upper bound on permissible penalties, where:

$$\overline{\Delta}_L < \frac{d}{p}, \qquad \overline{\Delta}_M = \frac{2d(p + q + (1 - q - p)^2)}{p}, \qquad \overline{\Delta}_H \geq \frac{d}{p(1 - q - p)} \tag{3.3}$$

The assumption of a discrete $\overline{\Delta}$ eliminates corner solutions. Note, however, that admission of a continuous $\overline{\Delta}$ does not change the qualitative findings of the model. $\overline{\Delta}$ can be interpreted as a measure of bureaucratic non-insulation. Lower values of $\overline{\Delta}$ constrain the punishment that the penalty the politician can impose, indicating a more insulated bureaucrat.

The citizen's decision about whether to complain depends on her valuation of the bureaucrat's allocation, and what can be recovered through the politician's auditing. I assume that citizens value receiving the service, regardless of state, i.e., citizens prefer to receive benefits, even when they are not "qualified." Citizens gain utility normalized to 1 if they ultimately receive the service and 0 otherwise. The citizen's utility function is given by Equation 3.4:

$$U_C(c; \theta) = \begin{cases} a - \theta c & \text{if } a = \omega \\ \rho(a, c)(1 - a) + (1 - \rho(a, c))a - \theta c & \text{if } a \neq \omega \end{cases} \tag{3.4}$$

The bureaucrat exerts effort in order to deter the penalty Δ. His utility is given by Equation 3.5, where r is an indicator taking the value of 1 if the politician audits and reverses the bureaucrat's allocation. $w > 0$ represents a wage that satisfies the bureaucrat's participation constraint.

$$U_B = w - \Delta r \tag{3.5}$$

The politician contracts the bureaucrat, specifying the probabilities of audit, $\rho(a, c)$, and sanction for errors, Δ. The politician seeks to optimize the accuracy of services provided to to serve to someone "like her," net the costs of investigation. Formally, I assume that politicians, like citizens, are indexed by type, denoted θ_P, and politicians maximize accuracy (congruence) to serve a citizen for whom $\theta = \theta_P$.

This specification of the politician's preferences represents the second major departure from Prendergast (2003). In a population in which citizens are differentiated, an additional assumption is needed to justify which citizens a politician seeks to serve. Consistent with a large literature on targeted distribution by politicians, the assumption here is simply politicians value the state's capacity to serve a specific type of citizen, not the population as a whole. The extension in Section 3.6 considers a politician that seeks to maximize the accuracy of service provision for serve all citizens. This assumption is closer in spirit to the welfare-maximizing principal in Prendergast (2003), but maintains the heterogeneity in citizen type. Given the present assumption that politicians maximize accuracy for their own type, the monitoring rates $\rho(a, c)$'s depend on whether a citizen of the politician's type will complain conditional on an allocation and state. Equation 3.6 gives the politician's expected utility. Each line of the equation corresponds to one state (ω). The politician's type, θ_P enters this calculation through anticipation of an identical citizen's complaint-making behavior.

$$E[U_P(\boldsymbol{\rho}(a,c),\Delta;\theta_P)] = \frac{1}{2}\left[(q+pe)(1-\frac{\rho(1,c)^2}{2}) + (1-q-pe)(\rho(0,c)-\frac{\rho(0,c)^2}{2})\right] + \frac{1}{2}\left[(q+pe)(1-\frac{\rho(0,c)^2}{2}) + (1-q-pe)(\rho(1,c)-\frac{\rho(1,c)^2}{2})\right] \tag{3.6}$$

3.2.1 Sequence, Equilibrium Concept

1. The politician chooses a contract specifying $\boldsymbol{\rho}(a,c)$ and Δ.
2. The state, ω is realized and revealed to only the citizen.
3. The bureaucrat chooses effort level, e, allocating the service to the citizen, a.
4. The citizen observes a and decides whether or not to complain, c.
5. The politician monitors according to the contract. When she monitors, any bureaucratic errors are reversed and the bureaucrat is punished.
6. Utilities are realized.

I characterize a Bayesian Nash equilibrium. The politician's contract is given by $\Delta \in [0,\overline{\Delta}]$ and $\rho(a,c) \in [0,1]$ for $a \in \{0,1\}$ and $c \in \{0,1\}$. The bureaucrat's effort is given by $e : [0,\overline{\Delta}] \times [0,1]^4 \to \{0,1\}$, and his allocation is given by $a : [0,\overline{\Delta}] \times [0,1]^4 \times \{0,1\} \to \{0,1\}$. The citizen's complaint strategy is given by the mapping: $c : [0,\overline{\Delta}] \times [0,1]^4 \times \{0,1\} \times \{0,1\} \to \{0,1\}$.

3.3 Equilibrium Analysis

Given that the politician pre-commits to the auditing strategy, consider first the citizen's decision to complain. First note that, if a politician audits, she will observe the state. As such, if the state were $\omega = 0$, no citizen would complain. Even if the citizen were allocated $a = 0$, they would not recover the service via an audit. As such, they will not complain. In contrast, when the state is $\omega = 1$, a citizen will complain if an increased probability of recovering the service exceeds the citizen's cost of complaint, θ:

$$\rho(0,1) - \rho(0,0) \geq \theta \tag{3.7}$$

This implies that there exists some threshold, $\widetilde{\theta} \equiv \rho(0,1) - \rho(0,0)$, above which citizens do not provide information to the politician via complaints. I refer to citizens for whom $\theta < \widetilde{\theta}$ as "legible" to the state. Building off of Scott (1998) and Lee and Zhang (2016), "legible" here refers to a citizen that could be induced to share private information about their state via a complaint to the politician. The informativeness of a complaint to the politician depends on both the citizen's type and the allocation. The the ability to complain provides information only if the citizen is legible and the bureaucrat allocates $a = 0$. If $a = 1$, the citizen has no incentive to complain in either state.

Lemma 3.1. ***Informational value of citizen (non-)complaints***:

(i) If $\theta > \widetilde{\theta}$, the citizen never complains $(c = 0)$. Upon observing no complaint, the politician's posterior belief about the probability of non-congruence is: $Pr(a \neq \omega) = 1 - q - pe$ for any a.

(ii) If $\theta \leq \widetilde{\theta}$, the citizen complains if and only if $\omega = 1$ and $a = 0$. As such, the politician's posterior belief about the probability of non-congruence between ω and a is:

$$Pr(a \neq \omega) = \begin{cases} 1 & \text{if } a = 0, c = 1 \\ 0 & \text{if } a = 0, c = 0 \\ 1 - q - pe & \text{if } a = 1 \end{cases}$$

Now, consider the bureaucrat's decision to allocate effort to an illegible citizen (of type $\theta > \widetilde{\theta}$). In this case, the bureaucrat's decision to allocate effort depends on the accuracy gains from effort (i.e. p) and the anticipated monitoring rate under either allocation, $\rho(1,0)$ and $\rho(0,0)$, respectively. It is incentive compatible to exert effort, $e = 1$, if:

$$w - \frac{1-q-p}{2}[\rho(1,0)+\rho(0,0)]\Delta - d \geq w - \frac{1-q}{2}[\rho(1,0)+\rho(0,0)]\Delta \tag{3.8}$$

$$\Delta \geq \frac{2d}{p[\rho(1,0)+\rho(0,0)]} \tag{3.9}$$

In contrast, the bureaucrat exerts effort to more serve a legible citizen (of type $\theta \leq \widetilde{\theta}$) if:

$$w - \frac{1-q-p}{2}[\rho(1,0)+\rho(0,1)]\Delta - d \geq w - \frac{1-q}{2}[\rho(1,0)+\rho(0,1)]\Delta \tag{3.10}$$

$$\Delta \geq \frac{2d}{p[\rho(1,0)+\rho(0,1)]} \tag{3.11}$$

Comparing the expressions in Equations 3.9 and 3.11, it is clear that if $\rho(0,1) \geq \rho(0,0)$, the bureaucrat can be induced to exert effort at a higher marginal cost, d, on behalf of a citizen is expected to complain when wrongly denied the allocation. In considering the bureaucrat's behavior, one further consideration is warranted. Namely, is it always incentive compatible for the bureaucrat to follow his investigation?

Given that the effort incentive, Δ, is symmetric for an error of either type, any incentive to allocate the service contrary to the findings of a bureaucrat's investigation must be driven by different monitoring rates. Suppose first that the bureaucrat's research suggests $\omega = 0$. If he distributes $a = 0$ but is wrong (with probability $1 - q - ep$), he draws a monitoring rate of $\rho(0,c)$ where c depends on citizen type (θ). The bureaucrat's expected utility from allocating a in line with his investigation is shown on the left-hand side of the following inequality. In contrast, if the bureaucrat goes against his research allocating $a = 1$, he is more likely to be wrong (with probability $q + ep$), but the monitoring rate, $\rho(1,c)$, may be lower as shown on the right hand side of the inequality.

$$w - (1 - q - ep)\rho(0,c)\Delta \geq w - (q + ep)\rho(1,c)\Delta \tag{3.12}$$
$$\frac{\rho(1,c)}{\rho(0,c)} \geq \frac{1 - q - ep}{q + ep} \tag{3.13}$$

Given the parametric assumptions, the right hand side of Equation 3.13 is bounded between 0 and 1. Therefore, if $\rho(1,c) \geq \rho(0,c)$, this condition is always satisfied and the bureaucrat will always follow his investigation. If this inequality does not hold, the bureaucrat will give $a = 1$ even when his research suggests that $\omega = 0$ to reduce the likelihood of an investigation. Consider now the case in which the bureaucrat's research suggests that $\omega = 1$. In order for the bureaucrat to allocate $a = 1$, the following inequality must hold:

$$w - (1 - q - ep)\rho(1,c)\Delta \geq w - (q + ep)\rho(0,c)\Delta \tag{3.14}$$
$$\frac{\rho(0,c)}{\rho(1,c)} \geq \frac{1 - q - ep}{q + ep} \tag{3.15}$$

Comparing Equations 3.13 and 3.15, it is clear that if all relevant monitoring rates are equivalent, the bureaucrat will always follow his investigation. One final observation is warranted: if $\Delta = 0$, the bureaucrat indifferent between ignoring and following his signal in all cases. I assume that bureaucrat's indifference is broken by following his investigation.

Finally, consider the politician's determination of the bureaucrat's contract. Recall that the politician is trying to maximize the probability that a citizen of type $\theta = \theta_P$ receives the "correct" service relative to their state, given costs of monitoring. The determination of the marginal legible citizen depends on the monitoring rates specified in the contract.

Consider first a politician that represents citizens that cannot be incentivized to complain regardless of the bureaucrat's allocation, $\theta_P > 1$. Because a citizen of the type $\theta = \theta_P$ will never complain, the politician never learns such a citizen's state. Substituting $\rho(0,0)$ and $\rho(1,0)$ into the politician's objective and maximizing, $\rho(0,0)^* = (1-q-ep)$, $\rho(0,1)^* = 0$, and $\rho(1,0)^* = (1-q-ep)$. Further, the politician always prefers that the bureaucrat exert effort to improve the accuracy of his allocation. To incentivize effort, the politician must set $\Delta \geq \frac{d}{p(1-p-q)}$. However, this is only possible when $\overline{\Delta} = \overline{\Delta}_H$. When $\overline{\Delta} \in \{\overline{\Delta}_L, \overline{\Delta}_M\}$, the politician cannot offer effort incentives. In either case, because $\rho^*(1,0) = \rho^*(0,0)$, the bureaucrat will always allocate a in line with her investigation.

Second, consider the case of a politician whose type can be induced to complain if wrongly denied the service, e.g., $\theta_P \leq 1$. Such a citizen will complain if $a = 0$ and $\omega = 1$, but not if $a = 0$ and $\omega = 0$. Substituting $\rho(1,0)$, $\rho(0,1)$, and $\rho(0,0)$ into the politician's objective and optimizing yields $\rho(0,0)^* = 0$, $\rho(0,1)^* = 1$, and $\rho(1,0)^* = (1-q-pe)$.

However, if the politician sets these monitoring rates, any bureaucrat for whom $q < 1$ would always accede to a citizen of type $\theta \leq 1$, by allocating $a = 1$ regardless of her investigation because $1 - q - pe < \frac{1-q-ep}{q+ep}$, the incentive compatibility constraint in Equation 3.13. This is a manifestation of the "truth-telling" problem identified by Prendergast (2003). However, the problem also manifests in a second form with a heterogeneous population of citizens. For a citizen of type $\theta > 1$, a bureaucrat will only face the prospect of monitoring when she allocates $a = 1$, since $\rho^*(0,0) = 0$. As such, for any $\Delta > 0$, the bureaucrat is better off allocating $a = 0$ to a citizen that would never complain. Because the politician's objective is to maximize service provision for someone "like her," the capacity loss from the former problem (acquiescence to a prospective complainant) is of concern, but the latter is not.

The politician can do strictly better than allowing the bureaucrat to accede to every prospective complainant. She can employ one of two strategies. First, she can eliminate incentives, setting $\Delta = 0$ and monitoring at these optimal rates. If the bureaucrat does not fear punishment, he will not accede to a citizen that would complain. Moreover, the bureaucrat then has no reason to uniformly deny a citizen that would not complain. However, when $\Delta = 0$, the bureaucrat cannot be induced to exert effort, reducing the accuracy of the allocation.

Alternatively, Equation 3.13 gives the highest highest ratio of monitoring rates under which a bureaucrat will not accede to a potential complainant. Maximizing subject to this incentive compatibility constraint $\frac{\rho(1,0)}{\rho(0,1)} = \frac{1-q-ep}{q+ep}$ reduces monitoring rates when a subject is denied the service to $\rho(0,1)^* = \frac{q+p}{q+p+(1-q-p)^2}$, and increases monitoring rates when the service is granted to $\rho(1,0)^* = \frac{1-q-p}{q+p+(1-q-p)^2}$. However, this second strategy is only available when a politician can offer effort incentives, if $\overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\}$. The politician faces a trade-off between less efficient monitoring and incentivizing bureaucratic effort. If the return to bureaucratic effort, p, is sufficiently high, providing incentives is preferred. Denote $\widehat{p}(q)$ as the solution to:

$$E[U_P(\rho(0,0) = 0, \rho(0,1) = 1, \rho(1,0) = 1-q, \Delta = 0)] =$$
$$E[U_P(\rho(0,0) = 0, \rho(0,1) = \frac{q+p}{q+p+(1-q-p)^2}, \rho(1,0) = \frac{1-q-p}{q+p+(1-q-p)^2}, \Delta \geq \Delta_M)],$$

expressed as a function of q. When $p \geq \widehat{p}(q)$ the politician prefers the contract with effort incentives even though she must monitor at a higher intensity. When $p < \widehat{p}(q)$, the politician prefers the incentive-free contract.

Inspection of the optimal $\rho(0,1)^*$ with and without incentives reveals that the composition of prospective complainants changes when incentives are used, as $\frac{q+p}{q+p+(1-q-p)^2} \leq 1$. This means that a politician of type $\theta_P \in (\frac{q+p}{q+p+(1-q-p)^2}, 1]$ can only incentivize complaints and bureaucratic effort by monitoring cases of complaints at a higher rate. To avoid forcing the bureaucrat to acede to a possible complainant, the politician must also increment the rate of monitoring when $a = 1$ $(\rho(1,0))$. The additional monitoring is costly to the politician and can only be sustained when re-turns to monitoring are sufficiently high. I denote the threshold at which the politician is indifferent to providing effort incentives as $\overline{p}(q)$, as in the previous case. When $p \geq \overline{p}(q)$, the politician opts for incentives; when $p < \overline{p}(q)$, the politician adopts the contract with information but no incentives ($\Delta = 0$). Note that because inducing effort costs more for a politician of this type, $\overline{p}(q) \geq \widehat{p}(q) \forall q$.

The equilibrium contracts are formalized in Proposition 3.1. In describing contracts throughout this paper, the subscripts indicate two qualitative features of the contracts – whether information transmission is incentivized (I) and whether bureaucratic effort incentives are provided (E).

Proposition 3.1. *A politician of type $\theta_P > 1$ implements the contract:*

(i) $\boldsymbol{\varrho}_{\emptyset} = \{\rho(0,0) = \rho(1,0) = 1-q, \rho(0,1) = 0, \rho(1,1) = 0, \Delta \leq \overline{\Delta}\}$ *if* $\overline{\Delta} \in \{\overline{\Delta}_L, \overline{\Delta}_M\}$

(ii) $\boldsymbol{\varrho}_E = \{\rho(0,0) = \rho(1,0) = 1-p-q, \rho(0,1) = 0, \rho(1,1) = 0, \Delta \in [\frac{d}{p(1-p-q)}, \overline{\Delta}]\}$ *if* $\overline{\Delta} = \overline{\Delta}_H$.

A politician of type $\theta_P \in (\frac{p+q}{q+p+(1-q-p)^2}, 1]$ *implements the contract:*

(i) $\boldsymbol{\varrho}_{\underline{IE}} = \{\rho(0,0) = 0, \rho(0,1) = \theta_P, \rho(1,0) = \frac{\theta_P(1-p-q)}{p+q}, \rho(1,1) = 0, \Delta \in [\frac{2d(p+q)}{p\theta_P}, \overline{\Delta}]\}$ *if* $\overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\}$ *and* $p > \overline{p}(q)$

(ii) $\boldsymbol{\varrho}_I = \{\rho(0,0) = 0, \rho(0,1) = 1, \rho(1,0) = (1-q)^2, \rho(1,1) = 0, \Delta = 0\}$ *else.*

A politician of type $\theta_P \leq \frac{p+q}{q+p+(1-q-p)^2}$ *implements the contract:*

(i) $\boldsymbol{\varrho}_{\overline{IE}} = \{\rho(0,0) = 0, \rho(0,1) = \frac{p+q}{p+q+(1-p-q)^2}, \rho(1,0) = \frac{1-p-q}{p+q+(1-p-q)^2}, \rho(1,1) = 0, \Delta \in [\frac{2d(p+q+(1-p-q)^2)}{p}, \overline{\Delta}]\}$ *if* $\overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\}$ *and* $p > \widehat{p}(q)$

(ii) $\boldsymbol{\varrho}_I = \{\rho(0,0) = 0, \rho(0,1) = 1, \rho(1,0) = (1-q)^2, \rho(1,1) = 0, \Delta = 0\}$ *else.*

(All proofs in appendix.)

Figure 3.2 depicts the qualitative features of the contracts in Proposition 3.1 graphically across the parameter space. The x-axis in all plots gives politician type, θ_P. The y-axis depicts the level of

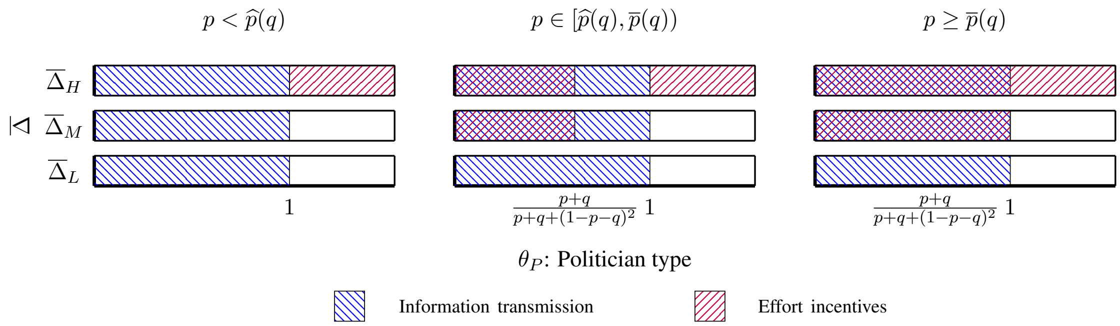


Figure 3.2: Visualization of the qualitative features of equilibrium contracts in Proposition 3.1 across the parameter space. The blue pattern shows that any politician for whom $\theta_P \leq 1$ will always incentivize information transmission from some citizens by setting $\rho(0,1) \geq \rho(0,0)$. The purple shading shows the regions of the parameter space where effort incentives are implemented, i.e. where Δ is large enough to sustain bureaucratic effort.

bureaucratic (non-)insulation, $\overline{\Delta} \in \{\overline{\Delta}_L, \overline{\Delta}_M, \overline{\Delta}_H\}$. The horizontal panels depict different levels of p, the accuracy gains attributable to bureaucratic effort. Three findings are of note. First, politicians of a low-cost type ($\theta_P \leq 1$) will always elicit information transmission from citizens. Second, the use of citizen information can support bureaucratic effort incentives at a lower $\overline{\Delta}$. Note that effort incentives are only possible at $\overline{\Delta} = \overline{\Delta}_M$ in the presence of information transmission. Finally, effort is not uniformly preferred in the presence of information transmission. This is a consequence of the trade-off between incentivizing bureaucratic effort and monitoring at optimal rates.

In the remainder of this paper, I use the contracts characterized here to develop implications for state capacity and inequality in societies of different compositions. The parameter determining the composition of a society is the density of θ and its cdf $F(\cdot)$.

3.4 Oversight and Implementation Capacity

I proceed by formalizing the definition of implementation capacity posited in the introduction. Specifically, implementation capacity is a measure of the state's ultimate ability to match service outputs to an unknown state associated with each citizen in a population. Given the definition of C as an indicator for the match between an allocation and the service provided, capacity is given by

$E[C]$, where the expectation is evaluated over both a citizen's state ω, and their type θ. As noted above, capacity need not be limited to the allocation of private goods (services) to individual citizens. The logic can productively be extended to the distribution of club or local public goods across communities as a function of communities' mobilization capacity.[4]

Definition 3.1. ***State Implementation Capacity:*** *State implementation capacity is the rate at which the ultimate service provided is matched to each citizen's state across the population, formally* $E[C]$.

As is clear from Definition 3.1, implementation capacity is not explicitly defined in terms of the amount of services given to a population or their distribution across the population, only the match between the allocation and a citizen's state. Because capacity is defined in terms of the ultimate service provided, the measure combines both the bureaucrat's equilbrium allocation and the politician's equilibrium monitoring strategy. As such, capacity incorporates bureaucratic effort, the bureaucrat's determination of whether to follow his investigation, and the rate at which the politician recovers the correct allocation via monitoring.

The contracts characterized in Proposition 3.1 also provide implications for the distribution of state services across the population. To this end, it is useful to examine $E[a^\dagger]$, the expectation of the ultimate allocation received by a citizen, as a measure of distributional outcomes.

Table 3.1 enumerates conditional expectations measuring capacity and distribution for different types of citizens. By conditioning on citizen type, this table clarifies several insights. The mapping between institutions (contracts) and the outcomes of interest – capacity and distribution – depends critically on societal composition. In particular, contracts that incentivize citizens to provide information (Contracts ϱ_I, $\varrho_{\overline{IE}}$, and $\varrho_{\underline{IE}}$) lead to different levels of capacity and allocation across the population, as a function of θ.

[4]This extension admits a reduced-form interpretation of θ as a community's mobilization capacity.

Contract	Citizen type, θ: $\theta \leq \frac{p+q}{p+q+(1-p-q)^2}$	$\theta \in (\frac{p+q}{p+q+(1-p-q)^2}, \theta_P]$	$\theta_P \in (\theta_P, 1]$	$\theta > 1$
PANEL A: ULTIMATE (POST-MONITORING) ALLOCATION $E[a^\dagger\|\theta]$				
$\boldsymbol{\varrho}_\emptyset$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\boldsymbol{\varrho}_E$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\boldsymbol{\varrho}_I$	$\frac{1}{2}(1+q-q^2)$	$\frac{1}{2}(1+q-q^2)$	$\frac{1}{2}(1+q-q^2)$	$\frac{1}{2}(2q-q^2)$
$\boldsymbol{\varrho}_{\overline{IE}}$	$\frac{2(p+q-pq)-q^2-p^2}{2(p+q+(1-p-q)^2)}$	–	0	0
$\boldsymbol{\varrho}_{\underline{IE}}$	$\frac{p+q+\theta[3(q+p)-2(q^2+p^2)-4qp-1]}{2(p+q)}$	$\frac{1}{2} + \frac{\theta[3(q+p)-2(q^2+p^2)-4qp-1]}{2(p+q)}$	0	0
PANEL B: CAPACITY BY CITIZEN TYPE $E[C\|\theta]$				
$\boldsymbol{\varrho}_\emptyset$	$1-q+q^2$	$1-q+q^2$	$1-q+q^2$	$1-q+q^2$
$\boldsymbol{\varrho}_E$	$q+p+(1-q-p)^2$	$q+p+(1-q-p)^2$	$q+p+(1-q-p)^2$	$q+p+(1-q-p)^2$
$\boldsymbol{\varrho}_I$	$\frac{1}{2}(2-q+q^2)$	$\frac{1}{2}(2-q+q^2)$	$\frac{1}{2}(2-q+q^2)$	$\frac{1}{2}(1+q^2)$
$\boldsymbol{\varrho}_{\overline{IE}}$	$q+p+\frac{1-p-q}{2(q+p+(1-p-q)^2)}$	–	$\frac{1}{2}$	$\frac{1}{2}$
$\boldsymbol{\varrho}_{\underline{IE}}$	$q+p+\frac{\theta_P(1-p-q)}{2(p+q)}$	$q+p+\frac{\theta_P(1-p-q)}{2(p+q)}$	$\frac{1}{2}$	$\frac{1}{2}$

Table 3.1: Implications of each contract for distribution, measured by $E[a^\dagger|\theta]$, and implementation capacity, measured by $E[C|\theta]$, by citizen type. The – indicates that this interval is empty when the contract $\boldsymbol{\varrho}_{\overline{IE}}$ is adopted.

One implication of Table 3.1 is that implementation capacity can only be achieved when bureaucrats can perfectly allocate the service, with or without effort.

Remark 3.1. *Bureaucratic quality and implementation capacity: Perfect bureaucratic quality, $q = 1$, is a sufficient condition to achieve complete implementation capacity, $E[C] = 1$. Complete implementation capacity cannot be achieved under any contract if bureaucratic capacity is incomplete, $q + p < 1$.*

When bureaucratic quality is perfect, $q = 1$, it is impossible (and unnecessary) to provide the bureaucrat with effort incentives. Either contract without effort incentives ($\boldsymbol{\varrho}_{\emptyset}$ or $\boldsymbol{\varrho}_I$) can be implemented in equilibrium. In this case, both yield observationally equivalent behavior since the bureaucrat never wrongly denies the benefit and the citizen never complains. It is possible to achieve complete capacity if $q + p = 1$ and the whole population was induced to complain under contracts $\boldsymbol{\varrho}_{\overline{IE}}$ or $\boldsymbol{\varrho}_{\underline{IE}}$. The remainder of the paper considers the remaining cases – arguably those consistent with empirical observation – in which bureaucratic capacity is limited ($q + p < 1$).

Consider the relationship between oversight institutions and implementation capacity. Figure 3.3 provides a visualization of state capacity under the contracts characterized in Proposition 3.1. The y-axis, $E[C|\theta]$ measures the likelihood that the ultimate (post-monitoring) allocation matches the citizen's state under each contract. Two findings are of note. In the absence of information, effort incentives weakly increase capacity. In contrast, the use of information from citizens increases capacity among those that can complain, while decreasing capacity among those that cannot.

As is evident from Figure 3.3, state capacity is calculated as the weighted average of $E[C|\theta]$ over the population type and state space. One immediate implication is that increases in capacity for legible populations generated by monitoring on the basis of complaints *reduce* the state's capacity to accurately serve the population that cannot complain. As such, using information transmitted

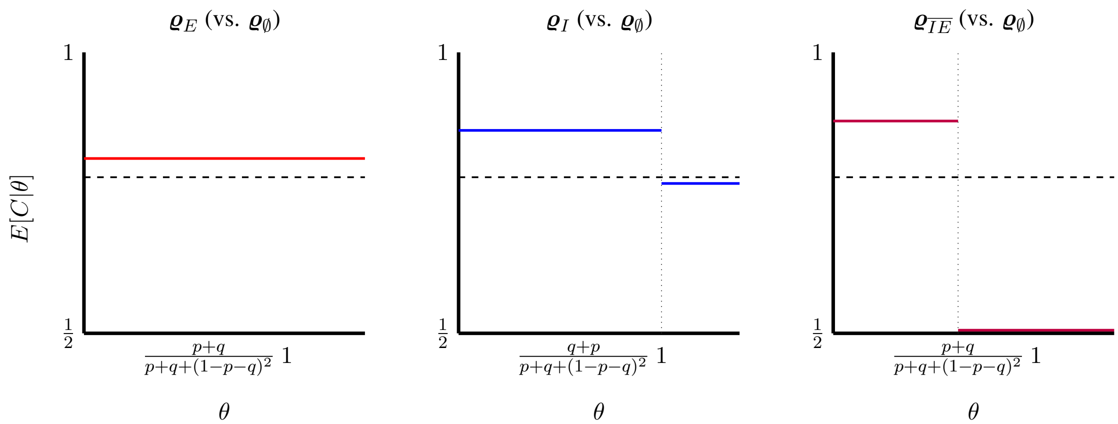


Figure 3.3: Graphical representation of state implementation capacity, conditional on citizen type, θ. The black dashed horizontal line measures capacity under contract $\varrho_\emptyset$, and serves as a benchmark. This benchmark is compared to the contract with effort incentives alone (ϱ_E) in the left panel; the contract with information transmission alone (ϱ_I) in the middle panel; and the contract with both features ($\varrho_{\overline{IE}}$) in the right panel. Note that contract $\varrho_{\underline{IE}}$ is substantively similar to the right panel.

by citizens in the form of complaints has an ambiguous effect on capacity depending on the proportion of citizens that can complain. As shown in Proposition 3.2, using information volunteered by citizens can only increase capacity when a sufficient share of the population can be induced to complain when wrongly denied the service.

Proposition 3.2. ***Information transmission and capacity****: Monitoring the basis of citizen complaints can increase or decrease state implementation capacity. Specifically, there exists a threshold, $\lambda \in [0,1]$, for which $F(\widetilde{\theta}) \geq \lambda$ implies that monitoring on the basis of citizen complaints weakly increases state capacity. If $F(\widetilde{\theta}) < \lambda$, monitoring on the basis of complaints decreases state capacity.*

Capture of the state by individuals or groups has been forwarded as corrosive to state capacity in different domains (Bardhan, 2002; Suryanarayan, 2020). Figure 3.3 suggests a novel mechanism underlying state capture. In contracts ϱ_I, $\varrho_{\overline{IE}}$, and $\varrho_{\underline{IE}}$, politicians commit to using information from citizens to monitor service providers. Such contracts, however, yield weakly less capacity to serve illegible citizens. The result resembles capture: by incentivizing some citizens to provide information to the state, legible citizens procure better service at the expense of the accuracy of services rendered to their illegible counterparts. The tradeoff between capture and the informational benefits of citizen complaints generates the ambiguous result in Proposition 3.2.

How do incentives to induce bureaucratic effort influence capacity? From Figure 3.3, it is clear that in the absence of information transmission, effort incentives increase capacity by increasing bureaucratic effort and accuracy. Specifically, moving from Contracts $\varrho_{\emptyset}$ to ϱ_E yields weakly higher capacity for all citizens, and thus the population as a whole. However, the effect of incentives is ambiguous in the presence of information transmission. Considering the difference between contracts $\varrho_{\overline{IE}}$ to ϱ_I, there are two counterveiling effects, as evidenced in Equation 3.16. Most obviously, effort incentives can increase the accuracy of targeting for at least citizens that complain. Less obviously, adding effort incentives reduces the share of legible citizens from $F(1)$ to $F(\frac{q+p}{p+q+(1-p-q)^2})$. Thus, while capacity is higher with incentives for a citizen of type $\theta < \frac{q+p}{p+q+(1-p-q)^2}$ in the parameter space in which $\varrho_{\overline{IE}}$ is adopted, these gains come at a cost of creating more illegible citizens and reducing capacity among these types. The comparison of $\varrho_{\underline{IE}}$ to ϱ_I yields a qualitatively similar finding.

$$E[C|\varrho_{\overline{IE}} - E[C|\varrho_I] = \underbrace{F\left(\frac{p+q}{p+q+(1-p-q)^2}\right)}_{\substack{\text{Legible under} \\ \varrho_I \text{ and } \varrho_{\overline{IE}}}}\left(p + \frac{3q-q^2-2}{2} + \frac{1-p-q}{p+q+(1-p-q)^2}\right) + \underbrace{\left(F(1) - F\left(\frac{p+q}{p+q+(1-p-q)^2}\right)\right.}_{\substack{\text{Legible under} \\ \varrho_I \text{ but not } \varrho_{\overline{IE}}}}\frac{-1+q-q^2}{2} + \underbrace{1-F(1)}_{\substack{\text{Illegible under} \\ \varrho_I \text{ and } \varrho_{\overline{IE}}}}\frac{-q^2}{2} \tag{3.16}$$

Note that in order for effort incentives to increase capacity when a principal employs monitoring on the basis of complaints, it must be the case that $F\left(\frac{p+q}{p+q+(1-p-q)^2}\right)$ is large and p is sufficiently high. In such cases, the composition of legible citizens is not substantially changed under the two contracts. Where differences in the set of citizens that complain under the two contracts are substantial, effort incentives lead to concentration of capacity among a smaller subset of the population at the expense of others, reducing overall capacity.

Politicians incentivize information transmission from citizens and effort incentives to increase the state's capacity to accurately serve a specific client. However, Proposition 3.2 establishes that the direction of the effect of these oversight tools on implementation capacity depends fundamentally on the underlying composition of the population as a whole. In particular, the use of information improves the state's capacity to serve legible citizens at the expense of other citizens. These dynamics are magnified in the presence of information incentives.

Collectively, the analysis of implementation capacity provides insights about the role of information transmission for bureaucratic oversight in comparative perspective. In particular, the effect of incentivizing citizens to volunteer information depends critically on the underlying legibility of the population. Where this distribution implies that an insufficient proportion can be made legible, relying on information from the legible population can harm outcomes for illegible citizens. The use of effort incentives alongside information transmission can compound these harms and can be particularly detrimental to state capacity in settings where few citizens are legible. This observation hints at the distributional consequences of oversight that are developed in Section 3.5.

3.5 Inequality

While oversight institutions have implications for capacity, they also influence the distribution of the service across a population. To this end, I proceed by considering the relationship between the types in a population – in terms of costs of complaint – and the distribution of state services. The service in question, a, is given to half the population if capacity were complete because $Pr(\omega = 1) = \frac{1}{2}$.[5]

[5] It is straightforward to see that the Gini coefficient on the capacity-maximizing allocation is equal to $\frac{1}{2}$.

However, because $\omega \perp \theta$, when the service is "perfectly" allocated, there are no differences in likelihood of receiving the service as a function of θ. The focus here is how inequality can emerge as a function of citizen type (cost of complaint), θ, under under the contracts enumerated in Proposition 3.1. As such, I develop a metric of inequality that abstracts from inequality generated by the state variable, ω.

The metric of inequality used to measure inequality as a function of θ is depicted geometrically in Figure 3.4. Specifically, I examine the share of total services devoted to each type of citizen. Note that under each of the contracts, there are at most two levels of service provision $E[a^\dagger|\theta]$, defined in terms of a (possible) cut-point, $\widetilde{\theta}$, which corresponds to the marginal legible citizen. Further, note that under any contract, moving from a citizen with $\theta = \theta'$ to a citizen of type $\theta = \theta''$ where $\theta' < \theta''$ implies that citizen of type $\theta = \theta'$ will, in expectation, be weakly more likely to receive the service. On the graph, the x-axis is the CDF of θ, $F(\cdot)$ and the y-axis is the cumulative share of service (a) received by citizens with lower θ's. The area of the shaded triangles thus visualizes the proposed metric of inequality, type-attributable inequality (TAI), defined formally in Definition 3.2. The maximum area of the triangle is theoretically $\frac{1}{2}$, so I normalize inequality to a more familiar $[0, 1]$ domain by doubling the area.

Definition 3.2. ***Type-attributable inequality (TAI)*** *measures inequality in the expectation of services provided as a function of citizen cost of complaint, θ. It is given by the formula:*

$$TAI(\varrho) = 2\mu_2\left((0,0),(F(\widetilde{\theta}),\frac{F(\widetilde{\theta})E[a^\dagger|\theta \leq \widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta \leq \widetilde{\theta}] + (1-F(\widetilde{\theta}))E[a^\dagger|\theta > \widetilde{\theta}]}),(1,1)\right)$$

where $\mu_2(\cdot)$ represents the area of the triangle defined by the three coordinates. $TAI \in [0, 1]$, and higher values of TAI indicate higher levels of inequality.

Proposition 3.3 describes the consequences of conditioning oversight on citizen complaints

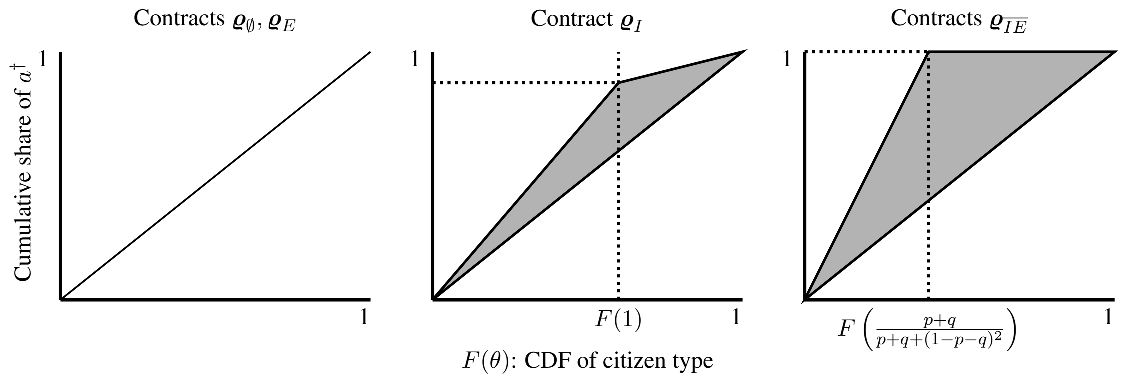


Figure 3.4: Geometric representation of inequality measure under each contract. The inequality measure, type-attributable quality (TAI), is equivalent to twice the area of the shaded region. Note that contract $\varrho_{\underline{IE}}$ is substantively similar to the right panel.

for inequality in access to services. As is evident from Table 3.1, when a contract precludes responsiveness to complaints, low-θ citizens respond by pooling with high-θ citizens by not lodging complaints, regardless of the service they receive. All citizens then receive the same allocation in expectation, resulting in no type-attributable inequality. In contrast, when politicians adopt monitoring systems that respond to citizen complaints, inequality in the expectation of service allocation emerges. Comparing the information-only contract (ϱ_I) to either of the contracts with information and effort incentives ($\varrho_{\overline{IE}}$ and $\varrho_{\underline{IE}}$), the combination of effort incentives and information generates weakly higher levels of inequality than information transmission alone.

Proposition 3.3. ***Oversight and inequality.*** *For any $q + p < 1$ and $F(1) \in (0, 1)$, conditioning oversight on citizen complaints introduces inequality in the allocation of a across the population, implying $TAI > 0$. TAI is weakly greater under contract $\varrho_{\overline{IE}}$ or $\varrho_{\underline{IE}}$ than under contract ϱ_I.*

Combining the discussion of capacity and inequality provides several insights about the implications of bureaucratic oversight. In societies in which some citizens can be induced to complain and others cannot, there exists a tradeoff between employing oversight institutions that maximize a state's capacity to define and serve legible citizens and equity in outputs. When examining capacity and equity across the whole population, using citizen information in oversight can either: (i) increase capacity at the cost of increased inequality in outputs; or (ii) decrease capacity while increasing inequality. The latter is more likely when relatively few citizens can be induced to complain (lower $F(1)$).

The canonical outcomes of civil service systems, captured here by higher bureaucratic quality (higher q) and higher bureaucratic insulation (lower $\overline{\Delta}$), reduce the magnitude of the inequalities that are generated by reliance on citizen complaints. However, this presents a paradox. States where bureaucratic quality is low or insulation is absent are low are precisely those places where use of citizen information in oversight can deliver the largest gains in capacity to (endogenously) legible citizens. As such, the distributional considerations highlighted here may be most salient in states with canonically weaker bureaucracies.

This model abstracts from considerations of political selection, or how θ_P is chosen. Given the importance of societal composition in determining the distributional consequences of oversight, some discussion of the the mapping between the distribution of politicians (the density of θ_P) and the population as a whole (θ) may shed light on the extent to which these dynamics are realized empirically. It seems plausible that potential politicians are disproportionately of a low θ_P relative to the population as a whole, whether in an autocracy or democracy.[6] If this is the case, we may expect a bias toward reliance on information transmission in oversight, relative to what might be generated by randomly drawing a politician from the population or an oversight strategy designed to appeal to the median voter.

[6] Arguments in support of the idea that politicians are drawn from a population with lower costs of complaint include: findings of positive electoral selection in democracies (e.g., Dal Bó et al., 2017; Besley and Reynal-Querol, 2011); accounts of dynastic politics and familial persistence in political office (e.g., Querubin, 2016; Smith, 2018); and studies of the importance of political connections for promotion in autocracies (e.g., Jia, Kudamatsu, and Seim, 2015).

3.6 Politician's Objective and Implications for Policy Design

To this point, I have considered a setting in which a politician maximizes the state's accuracy to serve a citizen "like her," net the costs of monitoring. The analysis thus considers the consequences of oversight policies based on an instance of an interaction between a single politician, a single bureaucrat, and a single citizen. The oversight strategy tailored to one citizen type is thus uniformly applied across to the population of citizens; in this case, the politician does not internalize the consequences of its application to other citizens. In this section, I consider different approaches to modeling a politician's design of oversight that considers the *population* of citizens.

I focus, therefore, on the oversight strategies adopted by a politician that maximize capacity across the population. In so doing, I depart from the assumption that monitoring rates are fixed across the population. In the previous analysis, I have assumed that the oversight contract is uniform across citizen types. Indeed, a politician catering to a particular type of citizen would set the monitoring rate at 0 for all other citizen types if she were allowed to do so. Bureaucrats would respond by not exerting effort for any citizen of type $\theta \neq \theta_P$, regardless of effort incentives, and the likelihood of receiving the correct allocation (ultimately) would fall to $E[C] = q$ across the population for any continuous distribution of θ.

Instead, consider a politician that seeks to maximize capacity (net of costs) for every citizen and can implement a contract as a function of citizen type. This approach aggregates across a population of citizens by simply considering infinite number of politician-bureaucrat-citizen interactions across the continuum of citizens, for which $\theta_P = \theta$ in each interaction. I assume that the bureaucratic quality and capacity are fixed across all interactions. This benchmark analysis thus abstracts from changes in the composition of who complains.

Following Proposition 3.1, thus, it is clear that for any $F(1) \in (0, 1)$ and $q < 1$, different monitoring rates will be adopted for different citizens. This can be visualized by examining Figure 3.2 for a fixed $\overline{\Delta}$ and p. Given that $F(0) = 0$ (by assumption), the politician adopts different contracts – here different monitoring rates – for different citizen types in the population. Importantly, every contract incentivizes information transmission from some citizens by setting $\rho(0,1) > \rho(0,0)$, as is the case in Contracts $\boldsymbol{\varrho}_I$, $\boldsymbol{\varrho}_{\overline{IE}}$, and $\boldsymbol{\varrho}_{\underline{IE}}$.

One implication of these capacity-maximizing contracts is that when there is variation in the legibility of a population, these contracts necessarily generate inequality in the allocation of the service across the population. This is evident from the variation in the levels of $E[a^{\dagger}|\theta]$, reported in Table 3.1. Thus, while information transmission from some citizens is necessary to maximize capacity across the population, when not all citizens can be induced to communicate, capacity is necessarily uneven.

Proposition 3.4. *Any capacity maximizing contract incentivizes information transmission from some citizen types. For any such contract, there exists inequality in expected allocations across the population ($TAI > 0$). However, relative to a uniform application of the most unequal "constituent" contract, conditioning the contract on citizen type reduces inequality (TAI).*

Proposition 3.4 further finds that the levels of inequality generated by the capacity-maximizing contract are *lower* than those generated by any of the constituent contracts with information when applied uniformly. This occurs because the use of type-specific monitoring rates effectively breaks the capture mechanism. The service provided to illegible citizens is no longer compromised due to oversight optimized for legible citizens. This implies that contracts that mandate unequal treatment of citizens by bureaucrats or their principals can *reduce* inequality in outputs. This finding has implications for a burgeoning literature on bureaucratic bias or discrimination. Studies that measure such biases often assert perverse implications (immediate or downstream) of differential treatment of citizens by bureaucrats ranging from disenfranchisement to inability to access state benefits (White, Nathan, and Faller, 2015; Hemker and Rink, 2017; **?**). The present result suggests that with a heterogeneous population of citizens, differential responses by bureaucrats to citizens of different types can actually reduce inequality in outcomes.

The finding that contracts that condition oversight on citizen type can reduce inequality in outputs raises several important considerations. First, benefits in terms of capacity or reductions in inequality that can be realized by conditioning oversight in this way rely on the assumptions about the politician's objective. If a politician is not motivated to, for example, maximize capacity for each citizen, the ability to monitor service provision differentially by citizen type can be corrosive to efforts to build capacity. This points to the importance of characterizing the agency problem between politicians and bureaucrats for the interpretation of empirical measures of bureaucratic behavior, particularly with respect to empirical documentation of bias or discrimination. Second, such conditioning may not be legal or feasible in certain contexts. Equal rights guarantees may preclude this form of conditioning state processes (here forms of oversight) on citizen type in program or institutional design.

This section extends the analysis to explore the tradeoffs between capacity and inequality when a politician seeks to maximize the accuracy of service provision across the population of citizens. One of the two central findings from the baseline model that monitoring bureaucrats on the basis of information transmitted by citizen complaints leads to inequality in service provision remains unchanged. However, in contrast to some cases in Proposition 3.2, using citizen information in monitoring cannot reduce overall implementation capacity across the population. Indeed, the politician is effectively maximizing implementation capacity when they opt to incentivize information transfer. Under this assumption about the politician's objective, therefore, there necessarily exists a tradeoff between state capacity (across the population) and equity that is not always present with a particularistic politician. The features of "capture" are eliminated by a politician that values accurate service provision across the population and can tailor oversight policies to the client. As such, the main results of this paper are robust to several specifications of the politician's objective. Nevertheless, accurate characterization of this objective can refine the predictions in terms of the effects of oversight.

3.7 Conclusion

This paper posits a new connection between oversight institutions and their implications for state implementation capacity and the distribution of state services across a population. Specifically, I examine how a politician's adoption of these institutions conditions the state's ability to accurately match service outcomes to an unknown citizen state. In particular, I study when oversight contracts provide citizens with incentives to complain about services rendered by bureaucrats. Citizen communication of private information helps the government match service to the citizen's state. However, when the state cannot incentivize all citizens to complain, a commitment to use information in monitoring improves capacity among legible citizens that can complain and can reduce implementation capacity among citizens that cannot. These dynamics generate inequality in the distribution of services across the population.

This theory speaks to many potential empirical applications. Most crucially, it emphasizes a broader role for the study of implementation in distributive politics. In the model, the policy is not targeted to any citizen type. It is directed on the basis of a citizen's state, which is assumed to be independent of observable characteristics (type). Yet, the politician's choice of contract generates substantial variation in "who gets what." In a large body of work that measures targeting of state resources in terms budgetary appropriations, the account of distribution developed in this paper would generally be undetectable in the data. As such, measuring only targeting in appropriations stage can yield inferences that do not reflect distributional outcomes.

Moving from theory to application suggests a need for new measures of the use of information transmission in bureaucratic oversight. First, more systematic data is needed to characterize variation in the role of citizen information provision cross-nationally and across policy areas subnationally. National regulation of procedures for complaint and information transfer appear to vary substantially in stylized cases. Moreover, measures of citizen willingness to provide information represent important variables in behavioral research that can develop our understanding of costs of complaint or legibility (as conceptualized in this article). Both sets of measures are critical to better formulating the relationship between politicians, bureaucrats, and citizens in the study of distributive politics.

This article views implementation capacity as the outcome of an interaction between a government and its subjects. By considering heterogeneity among citizens in terms of willingness to provide information, I provide a novel institutional foundation for observed unevenness in implementation capacity across the population or territory (e.g., Scott, 1998; Soifer, 2015). The model proceeds to link this unevenness to the co-occurrence of inequality in the distribution of state services. In so doing, it suggests new limits on states' ability to develop greater capacity for policy implementation without generating disparities in the distribution public goods and services.

Chapter 4

Bureaucratic Capacity and the Observability of Electoral Accountability

Electoral accountability is a normative goal of democracy (Przeworski, Stokes, and Manin, 1999). Yet, recent empirical assessments of the health or existence of these relationships between voters and politicians present grounds for pessimism. In developing democracies, widespread malfeasance by politicians, underprovision of public goods and services, and low levels of citizen political knowledge motivate questions about whether and when voters can hold these apparently poorly performing politicians to account. Furthermore, efforts to inform citizens in settings where accountability is thought to be limited provide little systematic evidence that citizens respond by sanctioning underperforming or rewarding high-performing politicians (Dunning et al., 2019).

In this paper, I present an alternative explanation for these empirical patterns that focuses instead on bureaucratic quality as a constraint on politicians' incentives to provide public goods, even when voters are sufficiently informed and rational. The theory suggests that empirical manifestations of accountability across democracies vary with bureaucratic quality. Specifically, I identify conditions under which the presence and absence of electoral accountability are observationally equivalent with respect to corruption by politicians, underprovision of public goods, and voter beliefs and behavior. These findings provide implications for how we interpret empirical evidence and design comparative research on accountability.

To see the distinction between my argument and existing alternatives, consider a standard account of accountability failures. In such models, citizens are uncertain about a politician's type and lack some informational signal upon which they can update their beliefs about the politician. This model is consistent with the implied equilibrium absent information in Ferraz and Finan (2008), Humphreys and Weinstein (2012), and Dunning et al. (2019), among others. Alternatively, citizens may receive information but fail to rationally update their beliefs (Achen and Bartels, 2016; Healy, Malhotra, and Mo, 2010).[1] In either case, failure to access or update on performance-relevant information breaks the link between a politician's actions and her subsequent electoral fortunes, so politicians maximize their own utility, regardless of whether their actions are congruent with voters' preferences.

In contrast, I argue that low bureaucratic quality generates the same empirical patterns even with voters that hold politicians to account. Across democracies, politicians rely on bureaucrats to co-produce public goods. Yet, politicians in different settings face bureaucrats of varying quality. I define quality as a function of human capital levels of bureaucrats and potentially their allocation across a politician's jurisdiction. Variation in bureaucratic quality influences the efficiency of public goods investments, altering politicians' incentives to appropriate funds to public goods. Citizens attempt to learn about the politician's type on the basis of observed public goods outputs, but the scope for learning varies with bureaucratic quality.

To understand the implications of these two accounts, I develop a simple two-period model of electoral accountability with a bureaucrat. As in standard models of accountability, voters evaluate politicians on the basis of the quality of observed policy outcomes, here levels of public goods pro-

[1] But see Fowler and Montagnes (2015); Fowler and Hall (2018); Ashworth, Bueno de Mesquita, and Fridenberg (2018).

vision (Ashworth, 2012). This model departs from standard practice by considering co-production of public goods by politicians and unelected bureaucrats. The politician allocates funds between public goods and private rents (corruption). In turn, bureaucrats produce public goods with the allocated funds. Bureaucratic quality influences the efficiency with which these public goods are produced. Politicians differ in their competence at "getting things done," which translates into better monitoring of bureaucrats. Voters prefer competent types and update their belief about a politician's type on the basis of public goods outputs. The voter then decides whether or not to retain the incumbent for a second period or to elect a challenger. I examine the manifestations of variation in accountability by allowing the probability that the voter observes the signal (public goods output) to vary.

I characterize the perfect Bayesian equilibria of the model. At low levels of bureaucratic quality, public goods provision is inefficient for both types of politicians, and a pooling equilibrium emerges in which neither type of politician invests in public goods. As a result, voter observation of outputs is uninformative, and even an voter that observes the signal cannot differentially retain competent types. At moderate levels of bureaucratic capacity, public goods provision is efficient for competent types but inefficient for incompetent types. In this case, voters observation of public goods allows for updating on politician type which yields higher retention of competent types and can induce incompetent type to make costly investments in public goods in their first term. Finally, at high levels of bureaucratic quality, a pooling equilibrium emerges in which both types make efficient investments in public goods. Again, in this equilibrium, voter observation of public goods outputs allows for updating which facilitates more frequent retention of competent politicians.

I illustrate the plausibility of the model by revisiting empirical evidence on the accountability of mayors in Brazil. Specifically, I develop and validate a measure of bureaucratic quality across Brazilian municipal bureaucracies. I use this measure to extend seminal studies of corruption by politicians (Ferraz and Finan, 2011; Avis, Ferraz, and Finan, 2018); voter updating on politician performance (Weitz-Shapiro and Winters, 2016*a*; Winters and Weitz-Shapiro, 2016); and voter sanctioning of incumbent politicians on the basis of information disclosure (Ferraz and Finan, 2008). I find that a separating equilibrium where competent invest in public goods and incompetent politicians invest in rents emerges only in municipalities with comparatively low measures of bureaucratic quality (relative to the random sample of Brazilian municipalities). In municipalities with higher quality bureaucracies, a pooling equilibrium emerges in which all politicians invest in public goods (in both periods). The evidence presented allows for rejection of cases of the model characteristic of existing empirical literature in which (i) voters are (completely) uninformed or unable to update; or (ii) bureaucratic co-production of public goods is not a feature of citizen-politician accountability relations.

This paper contributes to theoretical and empirical literatures. First, the model considers a strategic relationship between a citizen, a politician, and a bureaucrat. It connects to accountability models focused on a voter and politician(s) (Fearon, 1999; Ashworth, Bueno de Mesquita, and Fridenberg, 2017) and to models of moral hazard in bureaucracies. Joining Yazaki (2018); Li, Sasso, and Turner (2019), the model posits distinct but complementary roles of politicians and bureaucrats in the production of public goods as central to our understanding of accountability. This represents a departure from spatial models of delegation such as Fox and Jordan (2011) and analyses of optimal institutional design (e.g., Maskin and Tirole, 2004; Alesina and Tabellini, 2007) in which bureaucrats and politicians (ultimately) use the same instrument to affect policy. These departures from standard accountability models allow for novel results characterizing why patterns of accountability manifest differently in some democracies than others.

The results engage large empirical literature on information and accountability in developing democracies (Dunning et al., 2019; Chong et al., 2015; Banerjee et al., 2011; Bhandari, Larreguy, and Marshall, 2019; Cruz, Keefer, and Labonne, 2018). The findings of these studies are mixed with respect to the (average) effects or non-effects of information on voter beliefs and election outcomes (Enríquez et al., 2019). This study helps to rationalize mixed results in this literature and provides guidance for research design. Joining an important recent contribution by Martin and Raffler (2019), I argue the co-production of public goods by bureaucrats and politicians conditions voters learning about politicians and their behavior. In contrast to this work, however, I find that the argument that bureaucratic co-production of public goods hinders voter learning is not general across the parameter space; when bureaucratic quality is high, reliance on bureaucrats to produce public goods can actually *improve* voter information and thus political selection. Empirically, I provide evidence that this equilibrium presents in substantial number of Brazilian municipalities.

In both academic and policy research, much recent emphasis on improving accountability emphasizes interventions targeting voters. The most salient (or at least numerous) interventions examine how creating a better-informed citizenry can facilitate better selection of politicians or improve politician behavior in office. The theory in this article finds that such voter-focused interventions are only capable of improving accountability when bureaucratic at moderate levels of bureaucratic quality when voter information is limited. In contrast, interventions focused on improving governance from the inside – by building capacity – hold promise in a larger subset of cases.

4.1 Theory

Consider three actors: an incumbent politician, P, a bureaucrat, B, and a voter V. I study the production of public goods over two periods (terms). In each period, the politician and bureaucrat jointly produce public goods that are observed (or unobserved) by the voter. After the first term, there is an election in which the politician contests office against a challenger.

Politicians are of an incompetent or competent type, $\theta \in \{\underline{\theta}, \overline{\theta}\}$, respectively. The politician's type is private information to the politician and the bureaucrat. The voter holds a prior belief that the politician is a competent type with probability $Pr(\theta = \bar{\theta}) = \pi \in (0,1)$. I conceive of competence as ability to manage the bureaucracy or "get things done" via oversight.[2] Specifically, a competent politician monitors the bureaucrat at intensity $\overline{\theta}$ while an incompetent politician monitors the bureaucrat at intensity $\underline{\theta}$, where $0 < \underline{\theta} < \overline{\theta} < 1$.

Public goods are produced as a function of the funding allocated by the politician in period t and the quality and effort of the bureaucracy. Specifically, politicians allocate a budget, normalized to 1 in each period, between public goods (a_t) and private rents ($1 - a_t$). I assume a binary allocation decision, i.e. $a_t \in \{0, 1\}$.[3] I treat the quality of the bureaucracy, $q > 1$ as exogenous. While bureaucratic quality may be an outcome of policies pursued by a politician, the model simply assumes that quality is slow-moving and requires sustained investment to realize changes (Rauch, 1995; Huber and Ting, 2020).[4]

I assume that bureaucrat the bureaucrat exerts effort, e in response to some intensity of oversight, given by $\theta \in \{\underline{\theta}, \overline{\theta}\}$. As such, the utility of the bureaucrat, in period t, net of a wage satisfying his participation constraint, can be written:

$$u_t^B(e) = -\theta(1 - e_t) - \frac{e_t^2}{2} \tag{4.1}$$

Note that θ is given by the politician's type. The bureaucrat is myopic. Because the bureaucrat does

[2]There may be statutory regulations that constrain or empower the politician to take action. I abstract from these considerations at the moment, but they could be modeled as the product of θ and some variable capturing the statutory environment.

[3]Given that the politician's utility is linear in a_t, if $a_t \in [0, 1]$, the results are identical because the optimal allocation decision is always at a corner. However, additional out of equilibrium beliefs are necessary if the strategy space for a_t is continuous.

[4]There is less systematic empirical evidence on reforms designed to improve bureaucratic quality. Notably, however studied programs that are intended to improve bureaucratic quality via hiring (selection) are are initiated by higher levels of government from *outside* the localities they serve, not by local politicians who may be judged on the quality of their services (Dal Bó, Finan, and Rossi, 2013; Ashraf et al., 2020).

not internalize public goods outcomes, their utility is independent of their quality.

Given the allocation of funds by the politician and the effort exerted by a bureaucrat, the public good, $g_t(a_t, e_t)$ is produced according to the production function in Equation 4.2. The production function assumes that allocation to public goods and bureaucratic quality are complements.

$$g_t(a_t, e_t) = \begin{cases} a_t q & \text{with probability } e_t \\ 0 & \text{with probability } 1 - e_t \end{cases} \tag{4.2}$$

Equation 4.2 indicates that if the politician invests in public goods ($a_t = 1$), then the expected quantity of public goods outputs is increasing in bureaucratic quality (q) and effort (e_1). In contrast, if the politician starves public goods funding ($a_t = 0$), they are not produced. Equation 4.2 further clarifies the relationship between bureaucratic quality, q, and broader notions of bureaucratic capacity. As is discussed in existing work, bureaucratic capacity consists of both the skill of bureaucrats (Geddes, 1994), their allocation of bureaucrats across a jurisdiction (Acemoglu, García-Jimeno, and Robinson, 2015), and the effort exerted by bureaucrats. I capture the first two features in quality (q) and the third in bureaucratic effort (e_t). Thus, in the present framework both bureaucratic quality and effort increase the efficiency with which a politician's funding allocation is converted to a public goods output.

The politician trades off private rents for public goods when allocating the budget. Both types of politicians value the provision of public goods.[5] However, variation in the two types' efficacy in inducing bureaucrats to work is captured in the realization of g_t.

$$u_t^P(a_t; \theta) = 1 - a_t + g_t \tag{4.3}$$

[5]One can generate qualitatively similar results if politicians were distinguished by their objectives, i.e., a venal type that does not value public goods and an altruistic type that does. Since the emphasis here is on settings where public goods are not produced, I opt for a setting in which politicians do not vary in their preference for producing public goods.

The politician receives $u_t^P(a_t;\theta)$ for each period she is in office, and utility normalized to 0 if she is not in office. This normalization creates a re-election incentive for the politician. As such, the politician's utility over two periods is given by:

$$u^P(\mathbf{a};\theta) = \begin{cases} 2 - a_1 - a_2 + g_1 + g_2 & \text{if re-elected} \\ 1 - a_1 + g_1 & \text{if not re-elected} \end{cases} \tag{4.4}$$

The voter observes the realization of first-term public goods provision, g_1, and forms a posterior belief about the politician's type, $\mu(g_1)$.[6] To understand the role of voter information – here, whether or not a voter observes the public goods – in generating results, I assume that the voter observes this signal of first-term public goods with probability $p \in [0,1]$. While it is natural to think that incumbents, challengers, or civil society would publicize performance signals, the information and accountability literature generally assumes some some barrier to diffusion of this information, for instance, a lack of local media (Ferraz and Finan, 2008; Larreguy, Marshall, and Snyder Jr., 2020). The assumption of exogenous revelation solely maintains that these diffusion technologies are not manipulated by politicians in the short-run. Further, treating p as exogenous is consistent with the theoretical treatment of information revelation in experiments on information and accountability (Izzo, Dewan, and Wolton, 2019). In this paper, setting $p = 0$ allows for characterization of equilibria "without accountability," which I interpret as broadly consistent with assertions that voters are uninformed about politician performance or public goods outputs (see anecdotes throughout Dunning et al., 2019). As such, varying p allows for examination of the effect of increasing voter information, and ostensibly accountability pressures.

The voter values consumption of the public good, whether or not they observe it. This implies

[6] In this baseline model, the voter does not observe the politician's allocation or the bureaucrat's effort (a_1 or e_1, respectively).

that the voter cares about a politician's competence to the extent that competent politicians produce more public goods (in expectation). The voter's utility is thus given by expected public goods provision in the second period and a valence shock for the incumbent, parameterized as $\phi \sim U[-b, b]$, where $b > q$. The voter votes, $v \in \{i, c\}$, to re-elect the incumbent (i) or elect the challenger (c). If elected, a challenger acts as a first-period incumbent. For that reason, I index second-period actions by i and c, respectively. The voter's second period expected utility from the a vote for incumbent or a vote for a challenger (c), can be expressed:

$$E[u_2^V(i)] = E[g_2^i|\mu] + \phi \tag{4.5}$$

$$E[u_2^V(c)] = E[g_2^c|\pi] \tag{4.6}$$

I assume that if the challenger wins, she acts as a first-term incumbent. Thus, in evaluating $E[g_2^i|\mu]$ and $E[g_2^c|\pi]$, the voter considers differences in expected politician competence and differences in allocation behavior that depend on a politician's term.

4.1.1 Sequence and Equilibrium Concept

The game proceeds according to the sequence:

1. Nature determines θ, the incumbent's competence. Only the incumbent and bureaucrat observe θ.

2. The incumbent allocates a_1 to the public good.

3. The bureaucrat exerts effort e_1 to produce the first-term public good, g_1.

4. With probability p, the voter observes g_1 and forms a posterior belief about the politician's type, μ. The valence shock ϕ is revealed, and the voter chooses whether to re-elect the incumbent or elect the challenger.

5. If the incumbent was re-elected, she allocates a_2^i to the public good. Otherwise, the challenger allocates a_2^c to the public good.

6. If the incumbent was re-elected, the bureaucrat exerts effort e_2^i to produce the public good g_2^i. Otherwise, the bureaucrat exerts effort e_2^c to produce the public good g_2^c.

I characterize the Perfect Bayesian Equilibria (PBE) of the game. The incumbent's allocation decision is the choice $a_1 \in \{0, 1\}$. The bureaucrat's effort allocation is $e_1 \in \mathbb{R}_+$. Public goods production, $g_1 : \{0, 1\} \times \mathbb{R}_+ \to \{0, q\}$, maps the budget allocation and bureaucratic effort into a public goods output observed by all players. Voters update beliefs on the observation of public goods $\mu : \{0, q\} \to [0, 1]$ and the voter's voting strategy is a mapping $v : \{0, q\} \times [0, 1] \to \{i, c\}$. The second period incumbent's allocation strategy is a mapping $a_2^i : \{0, q\} \times [0, 1] \times \{i, c\} \to \{0, 1\}$. Finally, second period bureaucratic effort and public goods production represents the mapping: $e_2^i : \{0, q\} \times [0, 1] \times \{i, c\} \times \{0, 1\} \to \mathbb{R}_+$ and public goods provision represents the mapping $g_2^i : \{0, q\} \times [0, 1] \times \{i, c\} \times \{0, 1\} \times \mathbb{R}_+ \to \{0, q\}$. As in many signaling games, there exist multiple equilibria in some regions of the parameter space. I invoke the intuitive criterion refinement (Cho and Kreps, 1987). Under this refinement, the equilibrium characterized here is unique.

4.2 Equilibrium Analysis

First, consider the bureaucrat's equilibrium level of effort. By straightforward inspection of the bureaucrat's objective, it is clear that optimal effort, $e_t^* = \theta$. Clearly, the bureaucrat's effort depends only on the politician's type in either period. When combined with Equation 4.2, this optimal effort indicates that politician competence and bureaucratic effort are complements with respect to the production of public goods. The assumption contrasts with the idea that politician type (quality) and bureaucratic quality are substitutes, which is typically motivated by the observation that high-quality bureaucracies tend to insulate outputs from the follies of bad politicians. Instead, the model develops an alternate mechanism for this observed insulation focused on how bureaucratic quality shapes a politician's allocation.

Turning to the incumbent's second-term allocation strategy, the politician considers the expectation second-term public goods provision, $E[g_2^i(a_2^i, e_2^i)] = \theta q a_2^i$. Where $E[g_2^i(a_2^i, e_2^i)] \geq 1$, a politician will invest the budget in public goods, $a_2^i = 1$. In contrast, where $E[g_2^i(a_2^i, e_2^i)] < 1$, a politician will invest nothing, $a_2^i = 0$. The politician's optimal second-period allocation strategy is given by:

$$a_2^{i*} = \begin{cases} 1 & \text{if } q \geq \frac{1}{\theta} \\ 0 & \text{else} \end{cases} \tag{4.7}$$

Intuitively, if the bureaucracy is of sufficiently low capacity (low q), even the competent type has no incentive to fund public goods when it is inefficient to do so. This implies that even a competent politician that values public goods outputs will "take the money and run" when the state is incapable of efficiently producing public goods. On the other hand, when q is sufficiently high, both types will fund public goods. The efficiency gains in the provision of public goods from a high-quality bureaucracy thus induce both types of politician to fund public goods, insulating outputs (to some extent) from incompetent politicians.

Consider the voter's voting decision. The voter votes for the incumbent if $E[u_2^V(i)] > E[u_1^V(c)]$. Given the distribution of the valence shock, the incumbent's probability of victory, $\tau(\mu,\mathbf{a})$, is given by:

$$\tau(\mu, \mathbf{a}) = \frac{1}{2} + \frac{E[g_2^i|\mu] - E[g_2^c|\pi]}{2b} \tag{4.8}$$

$$= \frac{1}{2} + \frac{\mu E[g_2^i(a_2^i, e_2^i|\theta = \overline{\theta})] + (1-\mu)E[g_2^i(a_2^i, e_2^i|\theta = \underline{\theta})]}{2b} - \frac{\pi E[g_2^c(a_2^c, e_2^c|\theta = \overline{\theta})] + (1-\pi)E[g_2^c(a_2^c, e_2^c|\theta = \underline{\theta})]}{2b} \tag{4.9}$$

Turning to the voter's beliefs and voting decision, recall that the voter observes g_1 with probability p. With probability $1-p$, the voter does not observe g_1. Consider the latter case first. Consistent with descriptive accounts, it may be the case that voters do not observe g_1 due to lack of attention to or access to media. In this case, $\mu = \pi$, which follows (trivially) from Bayes' rule. If voters do not update, a politician's re-election fate is independent of her first-period allocation decision. As such, the politician maximizes her utility by adopting the same allocation strategy in both periods, always adopting the optimal allocation strategy given by Equation 4.7. Thus, following Equation 4.9, the probability of re-election is $\tau(\pi, \mathbf{a}) = \frac{1}{2}$.

In the case that voters do observe g_1, they are able to update their beliefs on the basis of observed outputs. However, at different levels of bureaucratic quality, the signal offered by the realization of public goods differs in its informativeness. Politicians choose their first period allocation behavior on the basis of the efficiency with which a public good could be produced combined with their anticipated prospects for re-election.

Proposition 4.1. ***Equilibrium*** *In the unique perfect Bayesian equilibrium:*

(i) If $q < \frac{1}{\bar{\theta}}$*, both types of politicians allocate* $a_1 = a_2 = 0$ *to public goods.*

(ii) If $q \in \left[\frac{1}{\bar{\theta}}, \frac{2b(1-\pi\bar{\theta})}{\underline{\theta}(2b(1-\pi\bar{\theta})+p\bar{\theta}(1-\pi))}\right)$*, a competent-type politician allocates* $a_1 = a_2 = 1$ *while a incompetent-type politician allocates* $a_1 = a_2 = 0$ *to public goods.*

(iii) If $q \in \left[\max\{\frac{1}{\bar{\theta}}, \frac{2b(1-\pi\bar{\theta})}{\underline{\theta}(2b(1-\pi\bar{\theta})+p\bar{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right)$*, a competent-type politician allocates* $a_1 = a_2 = 1$ *while an incompetent-type politician allocates* $a_1 = 1$ *and* $a_2 = 0$ *to public goods.*

(iv) If $q \geq \frac{1}{\underline{\theta}}$*, both types of politicians allocate* $a_1 = a_2 = 1$ *to public goods.*

Consider the four cases in Proposition 4.1. In the first case, $q < \frac{1}{\bar{\theta}}$, bureaucratic quality is sufficiently low that investing in the provision of public goods is inefficient for either type. As a result, no public goods are produced in the first period. If $p = 1$, the citizen observes no public goods and updates, but it must be the case that $\mu = \pi$, as an observation that $g_1 = 0$ provides no additional information about the type of politician. As such, the probability of re-election – even with informed voters – is $\frac{1}{2}$. Optimal allocation strategies, posterior beliefs, and re-election rates are therefore identical regardless of whether the voter observes public goods. This equilibrium is supported by a voter's off-path beliefs that, upon observation of a non-zero public goods output, the voter believes the politician to be of the competent type with certainty.[7] Note that the competent type of politician has no incentive to deviate by allocating funds to public goods in the first period because the voter has no additional incentive to retain a competent politician that will not allocate the budget to public goods in the second period. This rules out any equilibrium in which the competent type invests in public goods in only the first period.

In the second case, $q \in \left[\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\right)$. In this interval, the competent type of politician can provide public goods efficiently while the incompetent type cannot, as $E[g_1|\theta = \overline{\theta}] > 1$ and $E[g_1|\theta = \underline{\theta}] < 1$. Consider the voter's beliefs and voting strategy. When the voter observes g_1, they learn from its realization. If they observe $g_1 = q$, the voter knows that the politician is a competent type, $\mu = 1$. In contrast, upon observing $g_1 = 0$, the voter's posterior is $\mu = \frac{\pi-\pi\overline{\theta}}{1-\pi\overline{\theta}}$, which implies $\mu < \pi$. Combined with Equation 4.9, when a voter observes public goods outputs, competent type politicians are re-elected at a probability strictly greater than the incompetent type. In this interval, thus, accountability improves (from the perspective of voter welfare) the selection of politicians. Importantly, this parameter region can be empty. Many empirical studies start from an assumption of the existence of a separating equilibrium; the theoretical analysis suggests that this assumption may not obtain.

In the third case, when $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right)$, the incompetent type politician

[7] Note that these off-path beliefs eliminate equilibria in which neither politician contributes to the public good in the first period in other regions of the parameter space.

pools with the competent type politician in the first period, allocating her budget to the public good, but shirks in the second period by allocating no funds to public goods. Note that this equilibrium exists only when the voter (sometimes) observes public goods outputs. If the voter were never to observe outputs ($p = 0$), it is easy to verify that this interval would be empty. When $p > 0$, accountability pressures induce an incompetent incumbent to allocate funds to public goods in the first period when she would not otherwise do so in this interval. If the voter observes that the public good has materialized, they update their belief to $\mu = \frac{\pi\bar{\theta}}{\pi\bar{\theta}+(1-\pi)\underline{\theta}} > \pi$; if the public good does not materialize, the voter updates their belief to $\mu = \frac{\pi(1-\bar{\theta})}{\pi(1-\bar{\theta})+(1-\pi)(1-\underline{\theta})} < \pi$. Note that in this case, a voter prefers any competent politician to a first-period incompetent politician to a second-period incompetent politician. This generates two implications. First, competent-type politicians are re-elected at higher rates than incompetent-type politicians whenever $p > 0$. However, because of fears that a second-period incompetent-type will shirk, this case generates an incumbency *dis*advantage (Klašnja and Titunik, 2017).

In the final case, $q \geq \frac{1}{\underline{\theta}}$, both types of politicians allocate the entire budget to public goods in both periods, regardless of how observant the voter is (the value of p). However, it is more likely that a competent politician induces the bureaucracy to produce the public good given her investment. If the voter observes the public good, their posterior beliefs (under each realization of g_1) are identical to the previous case (with information). Consequently, the voter retains competent type politicians at a higher rate than incompetent type politicians. As a result, the likelihood that the second-period office holder is a competent type is higher than in the baseline case. Unlike the previous case in which a voter is concerned about second-period shirking by the politician, in this case she is indifferent between a first- and second-period politician, conditional on type.

As is standard in political accountability models, the inclusion of a voter that learns from observation of first-period public goods introduces two mechanism through which the equilibrium changes relative to model where the voters is uninformed (Fearon, 1999). First, the voter updates on politician type and their resultant voting strategy re-elects competent politicians at weakly higher rates than incompetent politicians. Second, these changes in the likelihood of re-election change a incompetent-type politician's first-period allocation strategy to fund public goods when they would not otherwise do so in some intervals of the parameter space. However, these manifestations of accountability do not manifest in every case of the equilibrium.

4.2.1 Observational Equivalence

The discussion of Proposition 4.1 suggests that the consequences of a more informed electorate vary across the equilibria identified. To examine this more systematically, I examine the conditions under which variation p – the likelihood that voters observe public goods outputs – manifests in observationally distinct equilibrium actions and beliefs. Proposition 4.2 identifies several observational equivalencies.

Proposition 4.2. *In the PBE characterized in Proposition 4.1:*

(i) If $q < \frac{1}{\bar{\theta}}$, the voter's posterior belief and therefore the probability of re-election of the incumbent are observationally equivalent for any $p \in [0, 1]$.

(ii) If $q \notin \left[\max\{\frac{1}{\bar{\theta}}, \frac{2b(1-\pi\bar{\theta})}{\underline{\theta}(2b(1-\pi\bar{\theta})+p\bar{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right)$, the optimal allocation strategy for an incumbent of either type is observationally equivalent for any $p \in [0, 1]$.

Proposition 4.2 posits that accountability only manifests in distinct voter beliefs, voter actions, and politician actions when bureaucratic quality is sufficiently high. In a large body on literature on information and accountability, a politician accrues rents from office at the expense of public goods because voters are not watching. Proposition 4.2 suggests that politicians that value public goods may accrue rents from office while voters watch because low bureaucratic quality makes investments in public goods inefficient. Voters cannot update on the basis of observing a lack of public goods outputs and are consequently indifferent between re-electing the incumbent or electing a challenger from the same pool of candidates.

In contrast, when bureaucratic quality and politician competence are high enough to render investment in public goods (by at least one type of politician) efficient, voters can update their beliefs by observing public goods outputs. This updating leads to a higher likelihood of retention of competent versus incompetent types. For some parts of the parameter space, re-election concerns induce incompetent-type politicians to make costly first-period allocations to the public good that does would not occur in equilibrium absent accountability pressures.

Collectively, this observational equivalence suggests that the empirical diagnostics used to assert an absence of electoral accountability in developing countries – corruption by politicians, underprovision of public goods, coexisting with "uninformed" voters – may not imply a lack of accountability. Moreover, by choosing different diagnostics of electoral accountability failures on the basis of equilibrium outcomes (e.g. corruption), we have less leverage to distinguish between these accounts.

4.2.2 Observable Politician Behavior

The model also provides implications about when the co-production of public goods by politicians and bureaucrats may hinder or facilitate selection or moral hazard of politicians (Martin and Raffler, 2019). To isolate the role of co-production, I compare a model in which the voter observes politician first-period allocation behavior (a_1) to the main model in which the voter observes public goods outputs (g_1). This model in which voters observe politician allocation behavior instead of public goods outcomes removes a source of randomness – whether or not public goods are executed by the bureaucrat.[8] As such, in a separating equilibrium, the politician's type is fully revealed to the voter.

[8] This is consistent with the characterization of corruption revelation treatments in Izzo, Dewan, and Wolton (2019).

Proposition 4.3. *In the unique Perfect Bayesian Equilibrium:*

(i) If $q < \frac{1}{\bar{\theta}}$*, both types of politicians allocate* $a_1 = a_2 = 0$ *to public goods.*

(ii) If $q \in \left[\frac{1}{\bar{\theta}}, \frac{2b}{\underline{\theta}(2b+p\bar{\theta})}\right)$*, a competent-type politician allocates* $a_1 = a_2 = 1$ *while a incompetent-type politician allocates* $a_1 = a_2 = 0$ *to public goods.*

(iii) If $q \in \left[\max\{\frac{1}{\bar{\theta}}, \frac{2b}{\underline{\theta}(2b+p\bar{\theta})}\}, \frac{1}{\underline{\theta}}\right)$*, a competent-type politician allocates* $a_1 = a_2 = 1$ *while an incompetent-type politician allocates* $a_1 = 1$ *and* $a_2 = 0$ *to public goods.*

(iv) If $q \geq \frac{1}{\underline{\theta}}$*, both types of politicians allocate* $a_1 = a_2 = 1$ *to public goods.*

The equilibria characterized in Proposition 4.3 are substantively similar to the equilibria characterized in 4.1. However, differences between the two results suggest that both the direction and magnitude of bureaucratic co-production of public goods on politician selection depend on bureaucratic quality. First, consider the effect of co-production on the moral hazard of the politician. Note that in the baseline model, the moral hazard of the politician can be overcome by accountability pressures at intermediate levels of bureaucratic quality, where public goods production efficient for only the competent type. Comparison of Propositions 4.1 and 4.3 demonstrates that the parameter space in which the incompetent type can be induced to pool with the competent type in the first period weakly expands when voters observe allocation behavior instead of outputs. In the baseline model in which voters may observe g_1, the bureaucrat's effort adds noise to the production of the public good. As such, the politician's type is not fully revealed to the voter in the instance that public goods are not realized. In contrast, in the present model in which voters may directly observe a politician's allocation, the voter learns the politician's type with certainty. This increase in voter learning creates stronger incentives for this first-period pooling.

Comparison of the models in which voters observe allocations as opposed to outcomes suggests that bureaucratic co-production of public goods can weaken or enhance the voters' ability to select competent politicians. Consistent with Martin and Raffler (2019), in a separating equilibrium in which competent and incompetent types make different allocations in the first period, the noise in the mapping from a politician's allocation to the public goods produced by bureaucrats reduces a voter's ability to select competent types upon observation of outputs as opposed to allocations. However, in contrast to the argument in Martin and Raffler (2019), in any equilibrium in which both types of politicians allocate funds to public goods in the first period, complementarities between politician competence and bureaucratic effort imply that voters can select better politicians when they observe outputs.

Relative to Proposition 4.1, Proposition 4.3 also reveals a broader set of observational equivalencies, as summarized in Proposition 4.4.

Proposition 4.4. *In the PBE characterized in Proposition 4.3:*

(i) For any $q < \frac{1}{\bar{\theta}}$ or $q \geq \max\{\frac{1}{\bar{\theta}}, \frac{2b}{\underline{\theta}(2b+p\bar{\theta})}\}$, the voter's posterior belief and therefore the incumbent's probability of re-election are observationally equivalent for any $p \in [0,1]$.

(ii) If $q \notin \left[\max\{\frac{1}{\bar{\theta}}, \frac{2b}{\underline{\theta}(2b+p\bar{\theta})}\}, \frac{1}{\underline{\theta}}\right)$, the optimal allocation strategy for an incumbent of either type is observationally equivalent for any $p \in [0,1]$.

In this model, the observational equivalence at low levels of bureaucratic quality is maintained. However, additional equivalencies emerge in the other pooling equilibrium when voters can observe politician behavior directly. The idea that voters' information comes from scouring municipal budgets to observe allocations seems less plausible than voter observation of public-facing public goods outputs. However, most partial-equilibrium empirical assessments of accountability focus on revelation of politician allocation or behavior (i.e. corruption) as opposed to public goods performance. For this reason, observational equivalence in both models is important for interpretation of evidence on accountability.

4.3 Revisiting Studies of Accountability in Brazil

I present empirical evidence in support of this theory of accountability relations between voters and politicians using empirical evidence from municipal governance in Brazil. To contextualize these findings, I extend the findings of a number of recent studies on information provision, voter updating, and mayoral corruption. The choice of Brazil allows me to leverage a new measure of bureaucratic quality to complement a rich existing literature. These practical and epistemological justifications should not undermine the importance of the study of accountability in one of world's largest democracies.

The parameter emphasized throughout the paper, bureaucratic quality, occupies a similarly salient role in the analysis. I provide original measurement of bureaucratic quality at the municipal level. However I do not manipulate bureaucratic quality, nor do I identify a research design which generates plausibly exogenous variation in this variable. Indeed, historical evidence and existing theories of bureaucratic reform emphasize the rarity of such moments in addition to the strategic and dynamic nature of reform (generally described in terms of civil service adoption) (Geddes, 1994; Ting et al., 2012; Huber and Ting, 2020). As such, I rely on observational variation in bureaucratic quality in a series of analyses intended to test several implications of the theory.

4.3.1 Measuring Bureaucratic Quality

As the theory clarifies, bureaucratic quality (q) is distinct from bureaucratic effort (e) and public goods outputs (g). As such, a measure of bureaucratic quality must abstract from these features. I thus operationalize quality as a measure of human capital of individuals employed in municipal administration. I rely on Brazil's Basic Municipal Information Survey (MUNIC) to measure characteristics of employment in (direct) municipal administration. This survey, implemented by the Instituto Brasiliero de Geografia e Estadística, requires muncipalities to report counts of public employees working in direct municipal administration, disaggregated according to several categories including education and contract type. Given that the raw data consists of counts of public employees, the level of cross-sectional aggregation is the municipality.

I operationalize bureaucratic quality as the average education level of bureaucrats working in municipal administration. The measure of bureaucratic quality capture features of a representative (average) public employee. I abstract from measures of the number of public employees per capita for two reasons. First, per-capita measures of municipal employment do not account for efficiencies of scale: running a fixed set of programs requires more employees per capita in small municipalities. Second, classic descriptions of patronage in Latin America include accounts of low-wage workers filling out the ranks of public employment (Calvo and Murillo, 2004; Grindle, 2012). This is generally believed to deter bureaucratic quality.

In treating bureaucratic education as a measure of quality, several legal, economic, and political considerations are warranted. Legally, municipal employees in direct administration should be hired with civil service provisions, though empirically adherence varies substantially, with many municipalities relying heavily on contractors. Variation in hiring practices is substantial across municipal governments, which accords with the wide observed variation in bureaucratic quality (e.g., Toral, 2019).

Average bureaucratic education clearly is driven, in part, by local labor market conditions. The scope of heterogeneity across Brazilian labor markets is likewise impressive. I account for regional variation using state fixed effects. I also use flexible covariate specifications to adjust for municipal population, average municipal education (years of education), formality (percent of workers working in the formal sector), and per-capita GDP. All of these features correlate with the measure of bureaucratic quality (see Figure C.3). However, they (collectively) account for less than 20% of the variation in bureaucratic quality, indicating that this variation is not simply a function of variation in local labor markets (see Figure C.4).

Analysis of persistence of this measure of bureaucratic education within municipalities over five waves of MUNIC (2005, 2008, 2011, 2014, and 2018) accords with some qualitative assumptions of the model. While quality is secularly increasing over time across the sample of (Figure C.2), within municipalities bureaucratic quality is sticky. The autocorrelation of bureaucratic quality between waves of the survey is 0.57, in approximately triennial surveys (Table C.2). I further show that such persistence obtains across the constituent education categories. In the model, q is treated as exogenous and the politician does not alter bureaucratic quality. The data is consistent with this assumption. Table C.3 reports the results of first-difference models that regress changes in bureaucratic quality (from consecutive waves of MUNIC) on indicators measuring a change in mayor and change in the mayor's party in an intervening election. I find no consistent evidence that changes in mayor or mayor's party yields differential shifts in average bureaucratic quality: point estimates are near-zero and precisely estimated. Further evidence from visualization of the ECDF of changes in bureaucratic quality shows no evidence of differential changes in variance (Figure C.5). This is consistent with the model's assumption that bureaucratic quality is approximately fixed (or slow-moving).

Finally, given the importance of separating the implications of the present theory of bureaucratic quality and accountability from theories premised on a lack of voter information, I examine the association between bureaucratic quality and local media presence in Figure C.6 and Table C.4. In the Brazilian municipal context, the presence of community radio is argued to be the most important form of media for diffusing local news (Ferraz and Finan, 2008; **?**; **?**). I show that while raw measures of bureaucratic quality and radio station presence are positively correlated, conditional on the local labor market covariates and state fixed effects, the conditional association between bureaucratic quality and radio presence is estimated to be a precise zero. Further, I utilize radio presence as an additional covariate to allay possible concerns.

4.3.2 Measuring Politician Allocations

In this analysis, I measure politicians' allocations to rents, $1 - a$, relying on the results of federal audits of municipal governments. Such audits have gained prominence across Latin America in the last two decades as a policy, and have been used to measure corruption in academic literature (Ferraz and Finan, 2008; Chong et al., 2015). In Brazil, the audits I use to measure allocations are conducted by the federal Controladoria-Geral da União (CGU), through a municipal auditing program inaugurated in 2003. Because municipalities in Brazil receive the majority of their budgets from the federal government, such audits cover sizable shares of municipal budgets.[9] Audits consist of visits by a team of federal officials to municipalities to oversee allocation and disbursement of funds and observe outputs. They report their findings in reports which are disseminated by local media (Ferraz and Finan, 2008).

Brazil's audits are oft-studied as a natural experiment because the federal government randomly selects municipalities by lottery. While this random assignment has facilitated studies of the effects of audits (i.e., Ferraz and Finan, 2008; Avis, Ferraz, and Finan, 2018), from the perspective of this paper, it primarily ensures random sampling of municipalities. This sampling ensures support across all levels of bureaucratic quality.[10] Support across the distribution of bureaucratic quality is essential to making inferences about the theory proposed by this paper, as I describe in more detail below.

A final consideration about the use of audit data to measure allocations considers whether audits measure the actions of politicians, here, mayors, or the municipal administration generally. I

[9]Among the sample of municipalities audited in the first rounds of randomized audits, audits covered 60% of local budgets.

[10]For the purposes of this paper, the ideal auditing selection would randomly sample at higher rates from the tails of the distribution of bureaucratic quality to increase statistical power.

follow existing studies of accountability in attributing corrupt or malfeasant spending to mayors. This is precisely the inference that experimental studies of accountability ask voters to make (Boas, Hidalgo, and Melo, 2019; Chong et al., 2015; Arias et al., 2019). In Brazilian municipalities, executives are responsible for proposing a budget and, alongside the city council, monitoring its execution (Gonçalves, 2013). It is not surprising, therefore, that opposition to audits in the form of lawsuits have come from elected politicians, not other municipal officials (Seabra, 2018). Furthermore, studies of the effects of audits find few consequences for bureaucrats, at least in terms of retaining employment (Ferrali and Kim, 2019).

4.3.3 Measuring Voter Updating and Behavior

I finally measure voter updating and voting behavior in response to provision of information about politician allocation behavior. The study of information revelation and voter updating has spawned a large body of recent survey experiments and experiments (for a list of these studies, see Incerti, 2019; Dunning et al., 2019; Bhandari, Larreguy, and Marshall, 2019). I extend a survey experiment fielded in Brazil as reported in Weitz-Shapiro and Winters (2016*a*); Winters and Weitz-Shapiro (2016) to measure how voter responses to evidence of politician corruption/lack of corruption vary in bureaucratic quality. This allows for the cleanest test of the voter updating mechanism in isolation, separated from the effect of a politician's (strategic) allocation behavior or baseline voter information.

I then turn to a re-analysis of the effects of audits on voting behavior inspired by (Ferraz and Finan, 2008), examining how the effect being (randomly) audited on vote share varies in bureaucratic quality. Recent scholarship emphasizes a learning effect stemming from the CGU audit program among politicians (Lichand, Lopes, and Medeiros, 2016; Avis, Ferraz, and Finan, 2018). These effects of audits are argued to stem from the threat of with legal sanction instead of electoral punishment. I abstract from these dynamics by focusing on the first rounds of audits preceding the 2004 municipal elections for measurement. This coincides with the period of highest auditing intensity (most municipalities audited per year) (Seabra, 2018).

In both designs, the manipulation to voter information, parameterized by p in the model, means that voters observe politician allocations (more specifically, corrupt spending), not public goods. By design, the mayor (or hypothetical mayor in the survey experiment) did not allocate funds in anticipation of this revelation. As such, both tests represent partial equilibrium tests of the model advanced here. The value added of these tests is to examine the plausibility of the accountability mechanism.

4.3.4 Mapping Theory and Research Design

Empirically, I focus on the beliefs and actions of two actors in the model: the politician and the voter, examining four predictions of the model. In assessing the plausibility of this model, I compare it to two less general cases which, I contend, capture arguments in the existing empirical literature:

1. *No bureaucratic co-production*: Assume that $\underline{\theta} = 0$ and $\overline{\theta} = 1$.[11] This eliminates both pooling equilibria and ensures that public goods production is independent of bureaucratic quality. If voters observe public goods, g, the incumbent's type is revealed. Per Fearon (1999), this is a model of pure selection.

2. *No electoral accountability*: Assume that $p = 0$. The voter does not observe a signal and thus does not update. This breaks the link between first-period allocations and electoral fortunes, implying that any politician that could be induced to allocate first-period funds to public goods in order to improve her electoral prospects if voters were informed would not do so.

The first case is obviously rather narrow. It refers to a distributive setting in which voters aim to select politicians who will not appropriate rents. This is certainly far from the only voter objective

[11] Note that this model falls outside parametric restrictions imposed on $\underline{\theta}$ and $\overline{\theta}$ in the model, so this is not precisely a "case" of the general model.

in theories of selection. However, it is consistent with the motivation of many studies based in low- and middle-income democracies including the studies from Brazil that I extend. Note further that cases #1 and #2 are not mutually exclusive. Indeed, the arguments in which bureaucrats are absent often coincide with findings of limited accountability.

To adjudicate the general ("unrestricted") case of the model from these alternatives, I motivate three reduced-form tests in Table 4.1. Given the measured parameters, the theory does not provide unambiguous directional predictions for a research design that exogenously manipulates voter information. Because studies of this design occupy a very prominent position in the literature, I examine the evidence on how information changes voting behavior inductively in light of the theory that I advance.

The tests I propose rely on observation of variation in equilibrium selection at different levels of bureaucratic quality. This presents a challenge because we do not know the true mapping between q and the empirical measure of bureaucratic quality. For example, it is possible that all municipalities have sufficiently uniform bureaucratic quality to ensure that all municipalities fall into one of the equilibria identified in Proposition 4.1. The inferences from the proposed tests rely upon variation in the underlying equilibria. To this end, I adopt three complementary approaches to interpretation of the data. First, description of the data and reading of the existing literature provides some guidance on what is plausible. Second, I leverage insights *across* the tests described in Table 4.1. Because these data come from different sources, qualitative consistency across tests is, in principle, more challenging to achieve and also more informative. Finally and most practically, I am careful to model bureaucratic quality flexibly in regression specifications given the possibility of non-linearities suggested by the model.

	Testable Implications			Empirical test
	"Unrestricted" model $p > 0, 0 < \underline{\theta} < \overline{\theta} < 1$	**No bureaucratic co-production** $\underline{\theta} = 0, \overline{\theta} = 1$	**No voter information** $p = 0$	
1	*Politician allocations to rents $(1 - a)$ weakly decrease in bureaucratic quality (q).*	Allocations to rents are independent of bureaucratic quality (q).	–	Original analysis using measure of rents from Avis, Ferraz, and Finan (2018).
2	*Politicians allocate more to rents in their second term $(t = 2)$ than their first term $(t = 1)$. This difference is attenuated to zero at low and high levels of bureaucratic quality.*	Politicians allocate more to rents in their first term than in their second term. This difference is independent of bureaucratic quality.	There is no difference in allocation to rents between first and second terms.	Extension of (Ferraz and Finan, 2011).
3	*At high levels of bureaucratic quality, a voter's posterior belief (μ) is equivalent to her prior (π) upon receiving a signal that a politician allocated no funds to rents $(a = 1)$.*	Voters update positively $(\mu > \pi)$ in response to a signal that a politician allocated no funds to rents at any level of bureaucratic quality.	Voters do not update $(\mu = \pi)$ in response to a signal of a clean politician at any levels of bureaucratic quality.*	Extension of survey experiment reported in Weitz-Shapiro and Winters (2016*a*) and Winters and Weitz-Shapiro (2016).

Table 4.1: Theoretical implications of the model (the "unrestricted" model) relative to the two cases thereof. Note that – indicates that the case does not make predictions that differ from the unrestricted model. *For the "no voter information" model, the interpretation in the text is that voters do not observe the signal; the prediction in the cell represents a slightly different interpretation that yields the same equilibrium. Formal motivation of these predictions is included in the Appendix.

4.3.5 Estimation

Table 4.1 suggests three implications of the theory that I proceed to test. In the first test, I examine the association between politician allocations to rents (corruption) and municipal bureaucratic quality. To examine this relationship, I estimate an OLS regression of the form:

$$Y_{msl} = \beta_0 + \beta_1 Q_m + \boldsymbol{\gamma}_s + \boldsymbol{\lambda}_l + \boldsymbol{\delta}\mathbf{X}_m + \epsilon_{msl} \tag{4.10}$$

where Y_{msl} is the proportion of audited funds allocated to rents in municipality m in state s, as measured in lottery round l. Q_m is the measure of municipal bureaucratic quality and β_1 is the coefficient of interest. $\boldsymbol{\gamma}_s$ is a vector of state fixed effects and $\boldsymbol{\lambda}_l$ is a vector of lottery round fixed effects. $\boldsymbol{X}_m$ is a matrix of decile indicators for each of four municipal-level covariates: population, average education, formality rate, and GDP per capita. The unrestricted model implies that $\beta_1 < 0$: corruption declines in bureaucratic quality. I also estimate this specification with tercile and quartile bins of Q_m given the potential for non-linearity suggested by the model.

Second, I examine whether term differences in the allocation of rents varies in bureaucratic quality. I test this prediction by estimating an OLS regression of the form:

$$Y_{msl} = \beta_0 + \beta_1 Q_m + \beta_2 \text{Second term}_m + \beta_3 Q_m \text{Second term}_m + \boldsymbol{\gamma}_s + \boldsymbol{\lambda}_l + \boldsymbol{\delta}\mathbf{X}_m + \epsilon_{msl} \tag{4.11}$$

where Y_{msl} is the share of audited funds allocated to rents in municipality m in state s, as measured in lottery round l. Q_m is the measure of municipal bureaucratic quality and Second term$_m$ captures whether the politician is in her second term in the 2001-2004 term (when the audits occurred). Bureaucratic quality, Q_m, is modeled linearly and in quantile bins in different specifications. $\boldsymbol{\gamma}_s$ is a vector of state fixed effects and λ_l is a vector of lottery round fixed effects. $\boldsymbol{X}_m$ includes the same set of flexible controls for local labor markets. The quantity of interest is $\widehat{\beta}_1 + \widehat{\beta}_3 Q_m$, the marginal effect of being a second term mayor, at a given level of bureaucratic quality. I also use an analogous regression discontinuity design to decompose selection from shirking effects, which I describe at greater length in Appendix C.4.

Finally, using survey experimental data, I estimate how the magnitude of voter updating in response to revelation of a clean mayor varies in bureaucratic quality. Given relevant arms of the experimental design of Winters and Weitz-Shapiro (2016), I estimate conditional average treatment effects (CATEs) of the clean and corrupt signals on updating, using OLS regressions of the form:

$$\begin{aligned} Y_{ims} =& \beta_0 + \beta_1 + \beta_2 \text{Clean signal}_i + \beta_3 \text{Clean signal}_i Q_m + \\ & \beta_4 \text{Corrupt signal}_i + \beta_5 \text{Corrupt signal}_i Q_m + \boldsymbol{\gamma}_s + \boldsymbol{\theta} \boldsymbol{X}_m + \epsilon_{ims} \end{aligned} \tag{4.12}$$

where Y_{ims} is the survey response of individual i in municipality θ in state s. The signal indicators measure the two treatment arms in the survey experiment. I cluster standard errors at the level at which Q_m is measured: the municipality. The estimator of the relevant CATE is $\widehat{\beta}_2 + \widehat{\beta}_3 Q_m$, which measures updating on the clean signal at different levels of bureaucratic quality. As in all models, I also consider binned quantile indicators of the bureaucratic quality measure.

4.4 Results

I present four findings. The first three examine the implications of the model elaborated in Table 4.1 to assess the plausibility of the theoretical model of accountability that I advance. The final result looks at voter behavior more inductively in light of the evidence amassed in the first three analyses.

4.4.1 Politician Allocations to Rents Decrease in Bureaucratic Quality

In a first observational test of the theory, I examine the relationship between bureaucratic quality and funds diverted to rents, as measured by municipal audits in 2003. Table 4.2, Panel A, presents the bivariate regression of the share of funds allocated to rents $(1 - a)$ to rents on the bureaucratic quality measure. A one standard deviation increase in bureaucratic quality (context) reduces rents by 1.4 percentage points or 22.5% of the sample mean. The conditional association of bureaucratic quality and mayors' allocations to rents remains substantively similar when including covariates. Given the right-skewed distribution of the share of corrupt spending, I also look at the logged outcome, with substantively similar findings in Columns 4-6. Consistent with the model, more flexible specifications of bureaucratic quality in Panels B and C reveal no evidence of non-monotonicity in this relationship. Corrupt spending is concentrated in the lowest quantiles of bureaucratic quality. On average, 7.3% of federal funds in the lowest tercile of municipalities and 7.5% of such funds in the lowest quartile of municipalities are spent in a corrupt manner. As shown by the tables, these rates drop in higher quantiles.

Further descriptive examination in Figure 4.1 of the corruption (rents) outcome suggests that the modal politician allocates was not detected to engage in corruption with the audited federal funds. Indeed, the median politician only allocates 1.9% of these funds to rents. This is consistent with a setting with circumscribed corruption. Per the model, this circumscription occurs when bureaucratic quality is high enough to induce at least one type of politician to invest in public goods (the separating equilibria or the pooling equilibrium where both types invest). I return to this observation in the subsequent analyses.

The finding that rent extraction declines in bureaucratic quality is consistent consistent with the general model advanced here, regardless of whether voters observe signals of an incumbent's performance. It is inconsistent with the special case of the model in which bureaucratic co-production of public goods is absent. I proceed to the subsequent tests to further adjudicate between these alternatives.

	Share of corrupt spending			Log(Share of corrupt spending + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
A. LINEAR BUREAUCRATIC QUALITY MEASURE (Z-SCORE)						
Bureaucratic quality	-0.014**	-0.014**	-0.017**	-0.012**	-0.012**	-0.014**
	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.006)
B. BUREAUCRATIC QUALITY MEASURE TERCILES (RELATIVE TO FIRST TERCILE)						
Bureaucratic Quality, Tercile 2	-0.009	-0.009	-0.009	-0.007	-0.007	-0.007
	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)
Bureaucratic Quality, Tercile 3	-0.027**	-0.026*	-0.036**	-0.023**	-0.022*	-0.029**
	(0.012)	(0.014)	(0.018)	(0.010)	(0.011)	(0.014)
B. BUREAUCRATIC QUALITY MEASURE QUARTILE (RELATIVE TO FIRST QUARTILE)						
Bureaucratic Quality, Quartile 2	-0.009	-0.003	-0.002	-0.006	-0.001	0.000
	(0.015)	(0.015)	(0.015)	(0.012)	(0.013)	(0.012)
Bureaucratic Quality, Quartile 3	-0.019	-0.021	-0.029*	-0.015	-0.018	-0.024*
	(0.015)	(0.014)	(0.015)	(0.012)	(0.012)	(0.013)
Bureaucratic Quality, Quartile 4	-0.029**	-0.030*	-0.042**	-0.025**	-0.025*	-0.034**
	(0.014)	(0.016)	(0.021)	(0.012)	(0.013)	(0.017)
State FE		✓	✓		✓	✓
Lottery FE		✓	✓		✓	✓
Demographic controls			✓			✓
Community indicator			✓			✓
Outcome Range	[0,0.794]	[0,0.794]	[0,0.794]	[0,0.584]	[0,0.584]	[0,0.584]
Outcome Mean	0.062	0.062	0.062	0.056	0.056	0.056
Outcome Std. Dev.	0.10	0.10	0.10	0.085	0.085	0.085
Num. obs.	448	448	448	448	448	448

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.2: Association between bureaucratic quality, q, and allocations to public goods, a. Funds diverted from public goods are measured as the share of corrupt spending, as defined by Avis, Ferraz, and Finan (2018). Heteroskedasticity-robust standard errors in parentheses.

4.4.2 Second-Term Shirking Disappears as Bureaucratic Quality Increases

Recall that in the model, the separating equilibrium in which politicians allocate money to public goods in the first period but not the second emerges under two conditions. First, intermediate bureaucratic quality makes public goods production by the incompetent type of politician inefficient while public goods production is efficient for the competent type. Second, voters are likely enough to observe the signal that costly diversion of funds to public goods in the first term can improve re-election prospects. Ferraz and Finan (2011) establish that corruption is, on average, higher among second-term mayors than first-term mayors. This contrasts directly with the predictions of the model

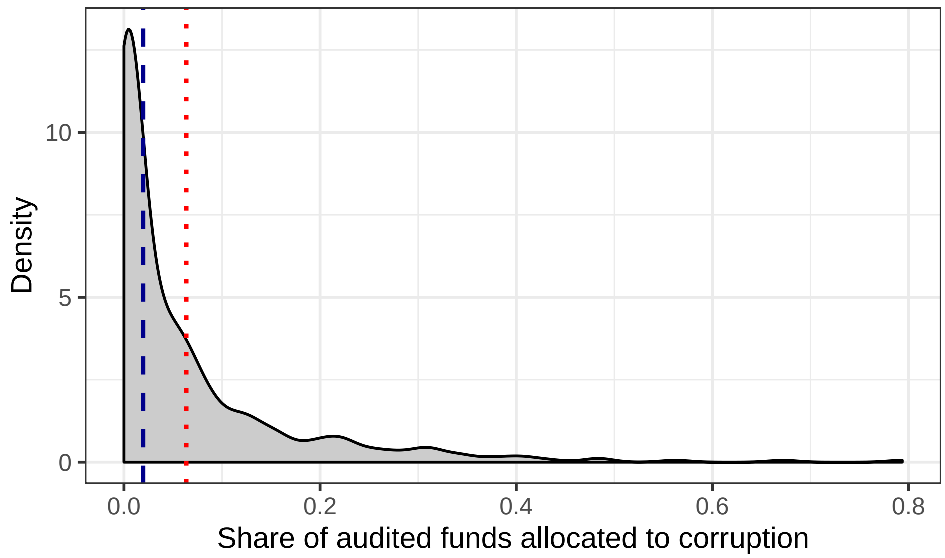


Figure 4.1: Distribution of the measure of rents: corrupt spending as a share of total audited federal funds. The navy dashed line gives the median and the red dotted line gives the mean.

without bureaucratic co-production, in which differences in allocations by term are driven only by positive selection in the re-election of mayors. I further examine how this difference between (term-limited) second term and (non-term limited) first term mayors varies in bureaucratic quality. Figure 4.2 shows that second-term shirking – or diversion of funds from public goods – manifests strongly in municipalities with low bureaucratic quality. However, this difference by term disappears as bureaucratic quality increases.

This finding is consistent with the suggestion that the lowest observed levels of bureaucratic quality correspond to the separating equilibrium and higher levels of bureaucratic quality correspond to the pooling equilibrium with public goods allocations from both types. Most importantly, under the model advanced here, the finding of any difference between first- and second-term mayors suggests that politicians anticipate that voters could learn about their actions (or outcomes). This provides evidence against the case of "no accountability" ($p = 0$). The result in Ferraz and Finan (2011) is sufficient for this conclusion. The finding that this difference varies predictably in

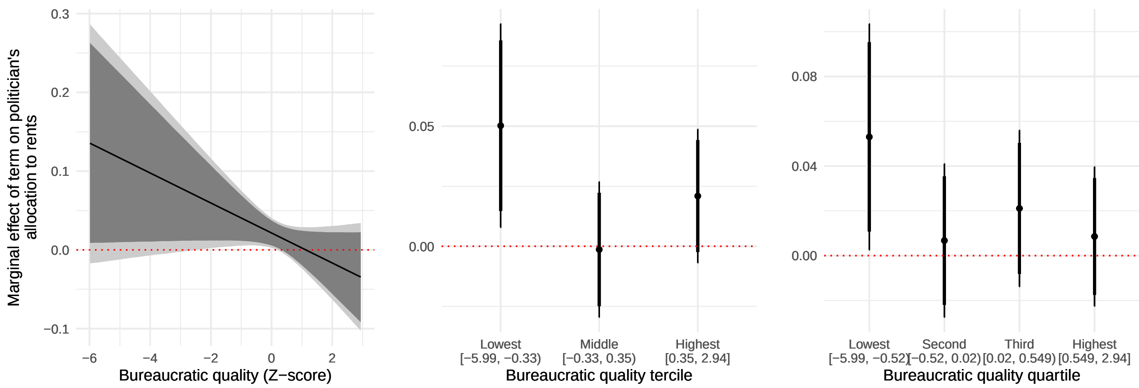


Figure 4.2: The marginal effect of a second-term politician versus a first-term politician on proportion of audited funds allocated to rents. 90% and 95% confidence intervals constructed on heteroskedasticity-robust standard errors.

bureaucratic quality provides evidence to bolster the plausibility of this model of accountability.

Serving as a first- versus second-term mayor is clearly not randomly assigned. One may be tempted to rely on a regression discontinuity (RD) design to identify the conditional (local) ATE of term.[12] Under the present model, however, an RD cannot identify the effect of term alone since second term mayors are also more likely to be of high competence than first-term mayors in the posited parameter space (consistent with the critique in Marshall, 2019). Yet, under the valence assumptions in the voting model, an RD-like estimator that narrows the bandwidth to close races provides a way to decompose term from politician type. By reducing the sample to close races (setting a small bandwidth on the RD), the share of incompetent types in the pool of second term incumbents should increase and we would expect more shirking on average (a larger difference between the first and second terms). By varying the bandwidth in Figure C.8, I show that this prediction is indeed borne out in the analysis. Consistent with the model's predictions, in samples where incompetent types theoretically represent a larger share of second-term mayors (close elections), second period shirking is more pronounced.

[12]Note that the difference in conditional LATEs would still not be causally identified without manipulation of bureaucratic quality.

4.4.3 Voters Do Not Update in Response to Revelation of "Clean" Politicians when Bureaucratic Quality is High

I extend a survey experiment to study voter updating in response to politician performance information at different levels of bureaucratic quality. In recent years, scholars have questioned the correspondence between voter (respondent) responses to information in survey experiments as opposed to field settings (Boas, Hidalgo, and Melo, 2019; Incerti, 2019). I contend that the two measure two measure beliefs and actions, respectively. Per the model, the two need not accord. I use the survey experiment to test whether voters update beliefs in a manner consistent with the theory. Replication and extension of the survey experiment fielded in 2013 and elaborated by Weitz-Shapiro and Winters (2016*a*) and Winters and Weitz-Shapiro (2016) thus affords a clean test of the updating mechanism in isolation.

The survey experiment provides respondents with a common vignette about a first-term mayor of a "different city" seeking re-election. A control condition provides no information about municipal audit outcomes. Additional text conveying a "clean" treatment condition indicates that the mayor was found to not have awarded bribes for city contracts. The "corrupt" treatment condition conveys that the mayor engaged in such bribery. Importantly, these three arms allow to test how citizens update in response to a signal of no allocation to rents (the "clean" condition) versus a signal of allocation to rents (the "corrupt" condition).[13] I measure the respondents' prior beliefs about the vignette via the control condition. I measure beliefs on a seven-point Likert scale.[14]

Importantly, the survey experiment was conducted with a nationally-representative survey in 140 municipalities. Because random selection of municipalities was weighted by population, the

[13]Boas, Hidalgo, and Melo (2019) replicate this survey experiment using nearly identical prompts in the state of Pernambuco during the 2016 municipal elections. However, the survey experimental portion of their study does not include a "clean" treatment condition, so I cannot test this implication in the context of their data.

[14]All findings are substantively similar when using Weitz-Shapiro and Winter's 2016*a* preferred "vote intention" outcome, as reported in Figure C.11. For the purposes of this analysis of updating, I seek a more direct measure of beliefs.

sample skews slightly toward higher levels of bureaucratic quality, though it maintains support over the distribution of quality, as shown in Figure C.9.[15] I estimate conditional average treatment effects (CATEs), by bureaucratic quality, as a measure of updating. The primary prediction of the model that I test holds that at high levels of bureaucratic quality, where competent and incompetent pool and allocate funds to public goods in a first term, voters should not update on the basis of a "clean" signal. The model does not offer predictions for the "corrupt" signal across the parameter space. At high levels of bureaucratic quality, diversion of funds from public goods is off the equilibrium path. The corrupt treatment condition thus allows for validation of the off-path belief assumptions in the PBE.

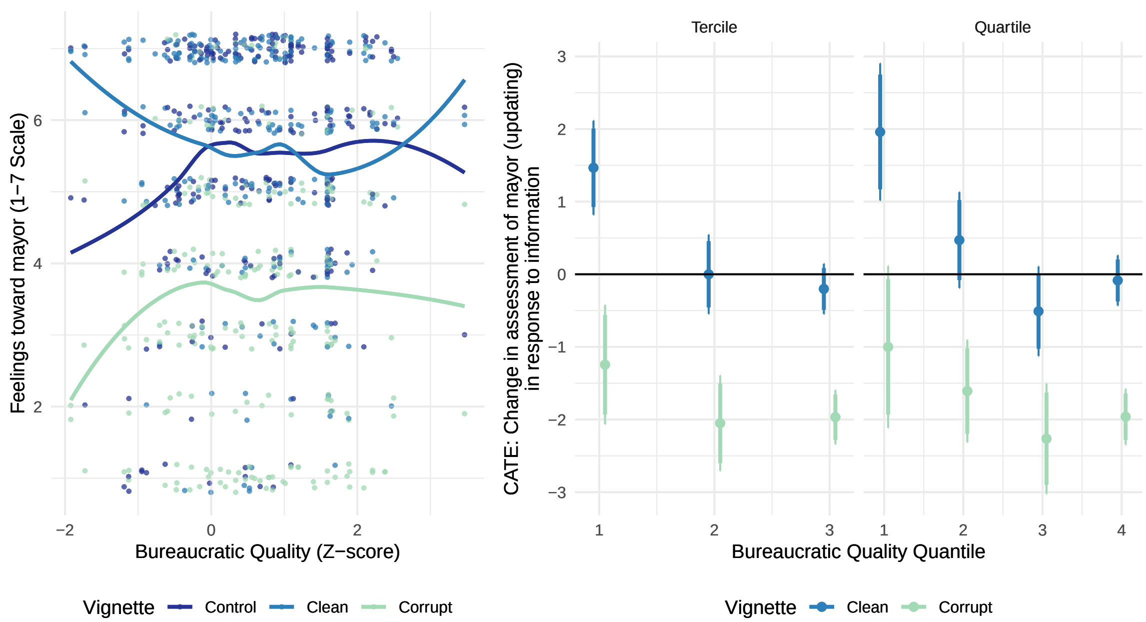


Figure 4.3: The left panel plots the raw data fitted by Loess for each treatment condition. The right panel estimates the CATEs of both "clean" and "corrupt" signals at different quantiles of bureaucratic quality. All CATEs are estimated from OLS regression models with the labor market covariates. Thick and thin lines represent 90% and 95% confidence intervals. Standard errors are clustered at the municipality, the level of aggregation at which bureaucratic quality moderator is measured.

[15] Because the survey was fielded in 2013, I use the 2011 levels of bureaucratic quality as the moderator of interest.

Figure 4.3 supports the model's prediction that voters update on the basis of a clean signal only at moderate levels of bureaucratic quality. In higher quantiles of bureaucratic quality, respondents' "average" prior (control) and posterior beliefs upon revelation of a clean mayor are not distinguishable.[16] Inspection of the data casts doubt on some alternative explanations. First, while the prior belief increases in bureaucratic quality, there is no evidence of a ceiling effect on the CATE, even high levels of bureaucratic quality, per the left panel of Figure 4.3. Across quantiles, there is "room" to observe positive updating on the basis of a clean signal. Second, while demand effects may be of concern with respect to voter updating about corruption, it is harder to explain why demand effects for the clean treatment would diminish with municipal bureaucratic quality. Finally, in this analysis, confounding is a concern with regard to the difference in CATEs at different levels of bureaucratic quality. However, this is not the relevant test of the theory; the CATEs are causally identified by the experimental research design (under standard assumptions). Thus, the relevant concern is that another feature that drives bureaucratic quality drives similar patterns of differential updating.

The strong treatment effects on corruption in lower quantiles of the distribution of bureaucratic quality are consistent with predictions of a separating equilibrium. At higher level of bureaucratic quality, the signal provided to voters is "off path" – such voters would not observe a signal about the politician's action in equilibrium. The negative updating observed in Figure 4.3 is consistent with the assumption that such a signal is generated by the incompetent type, though note that this is an assumption rather than a result.

I do not detect variation in these patterns of updating by individual respondent characteristics. In particular, these patterns do not vary detectably in citizen education or political knowledge on factual questions, as defined by Winters and Weitz-Shapiro (2016) (Figure C.12). This helps to ally

[16]The finding of substantial updating on politician in low quantiles is distinct from the null ATE across the sample that is reported in Winters and Weitz-Shapiro (2016).

concerns that differential patterns of updating are require sufficient political sophistication or are driven by variation in the educational composition of the electorate that covariates with bureaucratic quality.

In sum, extension of the survey experiment provides evidence that citizens (voters) respond to information by updating their beliefs in a manner consistent with the theory. Specifically, "good news" of a clean record politician only leads to updating when competent and incompetent types are predicted to take different actions. This suggests that voters internalize expectations about politician behavior and respond to information in light of these expectations. Per the model, differences in expectations are driven by bureaucratic quality.

4.4.4 Understanding the Effect of Information Revelation on Voting Behavior

A dominant research design in the study of accountability examines voter responses to the exogenous revelation of politician performance information. The randomized audits provide a natural experiment that researchers have used to study the effect of revelation of information on voters' decisions at the voting booth. A fundamental challenge in such research is measuring voters' prior beliefs, given that the direction and magnitude of possible updating depends on the this quantity. Existing approaches include pre-intervention surveys (i.e., Dunning et al., 2019) or more inductive approaches from the distribution of behavioral responses to information (Ferraz and Finan, 2008). In terms of the current model, testing the effect of information on voter behavior is complicated by incidental finding in the survey experiment that voters' prior appears to correlate with bureaucratic quality. This does not qualitatively change the general predictions of the model, but it limits the falsifiability of certain predictions related to the relationship between bureaucratic quality and voter behavior.

To show that voters appear to vote in a manner consistent with the model predictions, I rely on three pieces of evidence, which must be interpreted in tandem. The first evidence comes from

Ferraz and Finan (2008) who establish that voters, on average, punish incumbents found to engage in corruption when it is revealed before mayoral elections. I replicate this finding for municipalities audited before the 2004 elections using the metrics for corruption used in this paper in Table C.11. They further establish that this difference is driven by municipalities with local radio stations. I rely on these existing findings to provide evidence that voting behavior responds to the content of audits in the direction predicted by this theory. This is also consistent with the case without bureaucratic co-production, but is inconsistent with the model in which voters do not observe any signal of voter type ($p = 0$).

The second and third pieces of evidence rely on a distinct design that examines the full pool of mayors contesting for re-election in 2004 from municipalities that could have been selected for audits (those with a population under 450,000). One prediction of the theory holds that re-election rates should be lower in the equilibrium where incompetent types contribute to public goods in the first term by shirk in the second. Section 4.4.2 establishes that this occurs in lower quantiles of the observed distribution of bureaucratic quality. Observational analyses of re-election rates and changes in vote share as a function of bureaucratic quality provide relatively noisy evidence that incumbent electoral performance is weaker in lower quantiles of bureaucratic quality (Table C.12). For example, incumbents' re-election totals are 5.7 percentage points lower than their first-term vote share: on average, a 1 standard deviation increase in bureaucratic quality averts 0.7 percentage points of this shift, though the within-state comparison reveals a weaker association. As in the examination of term effects, this finding provides qualified support for the proposition that levels of bureaucratic quality influence equilibrium voter sanctioning behavior.

Finally, I examine the effect of being audited, unconditional of the information revealed, on incumbents' re-election performance using the natural experiment afforded by the random selection of municipalities into CGU's audits. First, I establish in Appendix C.13 that there is no evidence that contesting re-election varies with of being randomly selected to be audited. Indeed, nearly all eligible mayors contest re-election. Second, I show formally that when audits are randomly assigned, the prediction for this comparison is precisely zero at all levels of bureaucratic quality under the assumption that priors are an accurate measure of the proportion of competent and incompetent types in the candidate pool. This prediction provides one way to evaluate the rationality of voters' (behavioral) responses to information that remains agnostic as to (a) the (unmeasured) prior and (b) the (unmeasured) level of p. If voters did not update in a Bayesian manner, we would not expect an ATE of zero. Using equivalence tests, in Figure C.13, I reject standardized ATEs larger than 0.36 following recommendations by Hartman and Hidalgo (2018) in an omnibus test across all municipalities with an incumbent contesting re-election, and in general across quantiles of the distribution of bureaucratic quality.

Obviously, there are other rationales for observing a zero-effect of audits on vote shares and probabilities of re-election. Hence, these findings must be interpreted alongside the other results presented. If voters did not encounter the audits, we would similarly expect a zero effect; the results on directional updating in Ferraz and Finan (2008) are difficult to square with this explanation of the unconditional ATE of zero. Second, it may be the case that this quality dimension is simply swamped by other features when voters go to the polls (Boas, Hidalgo, and Melo, 2019). In the model, this captured by increasing the magnitude of the valence shock, b. If this were the case, we should also struggle to detect variation in electoral performance as a function of incumbent performance information revelation. In sum, the combination of evidence in this paper and predecessors does not offer evidence against the model I advance.

4.5 Beyond Brazil

The results in the previous section provide evidence consistent with the model advanced in this paper for the case of Brazil. To what extent can the model explain empirical patterns beyond this context? This is an open empirical question. In general, careful operationalization and measurement of variation in bureaucratic quality is necessary to extend the empirical examination to other contexts. While renewed efforts to measure bureaucratic quality at the national level are underway, the study of Brazil emphasizes the need to study variation in the quality of bureaucracies overseen by the politicians under study. Even without such data, however, the model offers some insights for how we consider evidence from other contexts.

Specifically, the selection of sites/cases poses underappreciated limitations for how we understand accountability and its failures. Drawing from the combined efforts of Enríquez et al. (2019) and Incerti (2019), I identify 16 experiments or natural experiments on voter information and accountability conducted in eight countries, enumerated in Table C.14. Situating these sites on (national level) macro indicators of bureaucratic quality, corruption, and public goods provision, Figure 4.4 suggests that such studies have been confined to democracies with low-to-middling levels of bureaucratic quality. Further, the motivation of some works suggests selection on features of the equilibrium, i.e. poor public goods provision or high corruption. This is also the case within country in many studies when units (i.e., municipalities) are sampled on the basis of the informational content of messages. Within a study, this non-random selection weakens our ability to leverage the effects of both "good" and "bad" news in tandem to measure the characteristics of an underlying equilibrium.

Examining eleven of these studies across all eight countries that: (i) are experimental and (ii) provide estimates on vote choice for the incumbent subsequent to the revelation of both good and

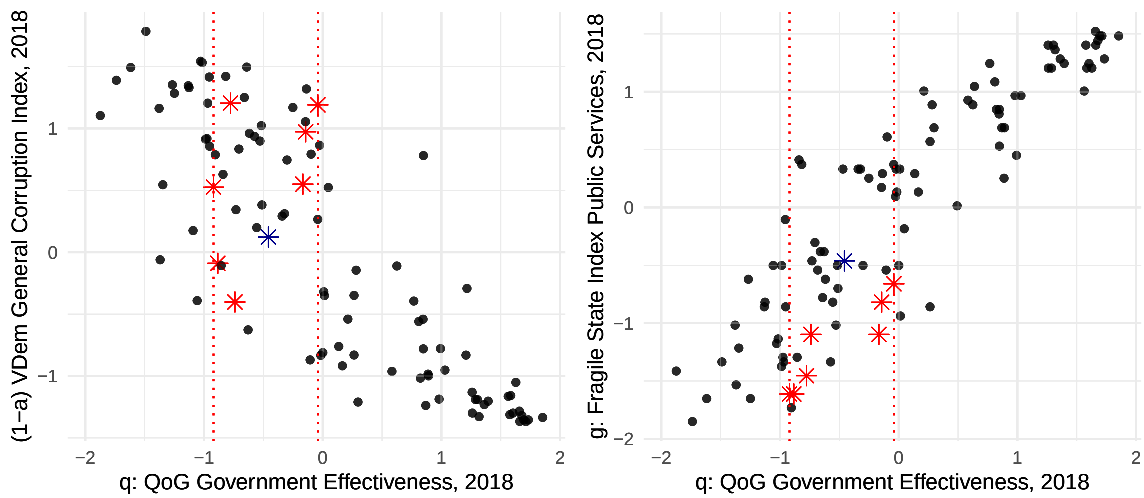


Figure 4.4: Correlations between bureaucratic quality (x-axes) and corruption (y-axis, left panel) and public service provision (y-axis, right panel). All measures are standardized to the set of democracies (defined by the Quality of Government dataset), such that all measures are z-scores. The navy star indicates Brazil and the red stars indicate other countries in which I identify accountability experiments/natural experiments.

bad news, Figure 4.5 depicts the relationship between national bureaucratic quality and the effects of information provision. On average, "good news" modestly increases incumbent vote share and "bad news" modestly reduces incumbent vote share, but only as bureaucratic quality (within sample) increases. Because all experiments are conducted in different constituencies the implication here is that the separating equilibrium appears to be more common (across constituencies) at higher levels of (national) bureaucratic quality. This finding is only suggestive and future research could strengthen this analysis in three ways. First, future experiments should avoid building the sample of constituencies by selecting on equilibrium outcomes (corruption or public goods provision). Second, measurement of the quality of bureaucrats actually managed by the politicians in question would allow for examining these dynamics much more precisely within cases. Finally, research designs that study accountability in a parallel fashion across places that with more substantial variation in bureaucratic quality (see Figure 4.4) would allow for more comprehensive tests of the theory

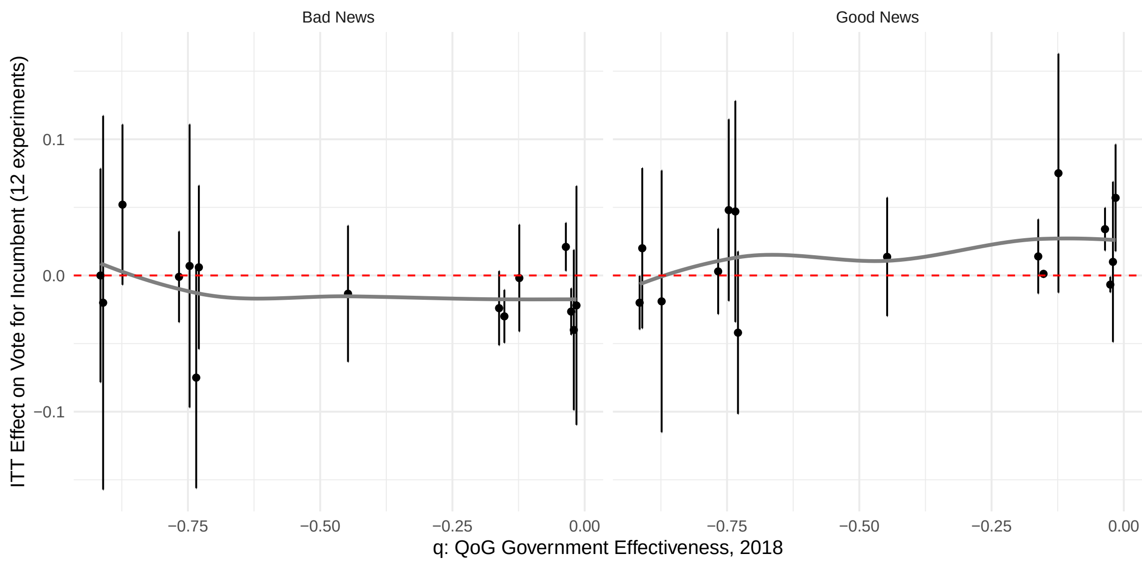


Figure 4.5: ITTs on vote choice for the incumbent or incumbent's party in 11 experiments in 8 countries using survey and/or administrative data. The dependent variable can be interpreted as the change in incumbent (party) vote share as a function of the revelation of incumbent performance information. Points are jittered on the x-axis for visibility. Estimates and standard errors come directly from estimated by these 11 studies.

than are afforded by the existing estimates.[17]

Considering this evidence from relatively similar studies across multiple sites yields two central takeaways. First, the observational equivalencies generated by the model provide new scope conditions on what inferences we can draw about accountability from partial equilibrium tests of information and accountability. Absent characterization of the underlying equilibrium, zero or null results on the effects of information provision provide less evidence that voters are uninformed or unable to update than is currently implied. Second, under the model I advance, estimates of a common distribution of treatment effects will be attenuated toward zero when sites fall into either pooling equilibrium. This is the central finding of two influential meta-analyses (Dunning et al., 2019; Incerti, 2019). Figure 4.5 provides suggestive evidence as to why such an approach may

[17]Figure 4.5 ultimately includes only eight country-level estimates of bureaucratic quality, which approximate "clusters" for the purpose of this analysis.

not be able to capture the effect of information on accountability. More generally, this reasoning suggests a need to define external validity relative to a theory (or argument), as opposed to a point estimate. This broader conception of external validity opens new avenues for cumulation of evidence and research design.

4.6 Conclusion

Recent literature on accountability remains mixed on whether a better informed electorate can improve political selection, drawing a (relatively) pessimistic conclusion that empowering voters with information is, at best, circumscribed in its ability facilitate better governance. The accumulated evidence tends to describe such outcomes in terms of the deleterious effects of uninformed voters and the venality of politicians. I contend that this attribution of bad outcomes – corruption and limited public goods provision – to a form of "bad politics" is particularly widespread in the study of developing democracies.

The theory advanced in this paper takes a different approach to explaining similar patterns of outcomes. Indeed, the model of electoral accountability advanced in this paper assumes that politicians uniformly value the provision of public goods. Voters are informed and rational (Bayesian). Bureaucrats shirk, but are responsive to oversight and are not otherwise corrupt. Yet, the theory predicts the confluence of corruption, underprovision of public goods, and voter behavior often taken to motivate claims of circumscribed electoral accountability can emerge with informed, rational voters when bureaucratic capacity is low. Furthermore, the dissemination of politician performance metrics conveys no information in the pooling equilibria at very low and high levels of bureaucratic quality. Extension of multiple existing results from Brazil emphasizes the plausibility of this model, providing grounds for extension and further testing in other contexts. Examination of treatment effects from twelve information dissemination experiments provides suggestive evidence consistent with the theory in a wider variety of contexts.

This paper advances a broader appeal for theories that treat political actors more symmetrically across contexts in order to generate comparative insight. To extent that the policy implications of empirical findings depend on the underlying causal process generating outcomes, a bias toward "bad politics" may limit the insight that we can contribute in efforts to inform policies to advance welfare.

Part I

Appendices

Appendix A

Appendix to Chapter 2

A.1 Theoretical Model: Extensions and Proofs

A.1.1 Proofs of Propositions in the Main Text

Proof of Proposition 2.1: $\mathbb{E}[e_g^*]$ is the expectation of bureaucratic effort for citizens of group g. Recall that c_C is a realization of the random variable C_g with the cdf $F_g(\cdot)$. As such, the proportion of citizens that would complain if service is not granted is given by $F_g(\frac{b}{c_P})$. Assume that effort is interior for all citizens, $e_g^* < 1$. It is straightforward to calculate $\mathbb{E}[e_g^*]$ from Equation 7:

$$\mathbb{E}[e_g^*] = \frac{\mathbb{E}[\gamma_B^g]}{c_B} + \frac{r}{c_B c_P}\left[S + \mathbb{E}[\gamma_P^g] + F_g\left(\frac{b}{c_P}\right)\right] \tag{A.1}$$

where $\mathbb{E}[\gamma_B^g]$ and $\mathbb{E}[\gamma_P^g]$ are the expectations of the bureaucrat's and politician's tastes for group g, respectively. The expectation of bias between groups x and y in the aggregate is defined as $\Delta = \mathbb{E}[e_x^*] - \mathbb{E}[e_y^*]$ and is calculated:

$$\Delta = \frac{\mathbb{E}[\gamma_B^x] - \mathbb{E}[\gamma_B^y]}{c_B} + \frac{r}{c_B c_P}\left[\mathbb{E}[\gamma_P^x] - \mathbb{E}[\gamma_P^y] + F_x\left(\frac{b}{c_P}\right) - F_y\left(\frac{b}{c_P}\right)\right] \quad \blacksquare \tag{A.2}$$

Proof of Proposition 2.2: Without loss of generality, assume that $\eta_S + \eta_Q > 0$. This implies that per citizen, the politician would devote more effort to monitoring the bureaucrat's service to group x when service is not provided. Note that if $\eta_B \geq 0$, the higher probability of oversight must increase bias. I solve for the magnitude of bureaucrat's taste-driven bias at which the magnitude (absolute value) of bias with and without oversight is equivalent. Comparison of the magnitude implies:

$$\frac{\eta_B}{c_B} + \frac{r(\eta_P + \eta_Q)}{c_B c_P} = -\frac{\eta_B}{c_B} \tag{A.3}$$

Simplifying:

$$\frac{r(\eta_P + \eta_Q)}{2c_P} = -\eta_B \tag{A.4}$$

This implies that for $\frac{r(\eta_P+\eta_Q)}{2c_P} > -\eta_B$, the magnitude of bias in effort is increased by the biased probability of oversight toward group x. For $\frac{r(\eta_P+\eta_Q)}{2c_P} = -\eta_B$, the magnitude of bias is not affected by oversight, and only when $\frac{r(\eta_P+\eta_Q)}{2c_P} < -\eta_B$, does the threat of oversight decrease bias. ■

Proof of Proposition 2.3: Comparative statics.

1. Differentiating Δ_O and Δ_B with respect to c_P yields:

$$\frac{\partial \Delta_B}{\partial c_P} = 0 \tag{A.5}$$

$$\frac{\partial \Delta_O}{\partial c_P} = \frac{r}{c_B} \frac{b(f_y(\frac{b}{c_P}) - f_x(\frac{b}{c_P})) - c_P(\eta_P + F_x(\frac{b}{c_P}) - F_y(\frac{b}{c_P}))}{c_P^3} \tag{A.6}$$

The sign of $\frac{\partial \Delta_O}{\partial c_P}$ depends on the shape of the densities $f_x(\cdot)$ and $f_y(\cdot)$. However, Δ_B does not vary in c_P while Δ_O may. Thus, $\frac{\partial \Delta}{\partial c_P} \neq 0$ indicates that $\Delta_O \neq 0$.

Further, note that $F_x(\frac{b}{c_B}) - F_y(\frac{b}{c_B}) \in [0, 1]$, by the assumption $F_y(\cdot)$ FOSD $F_x(\cdot)$. This implies that $\eta_Q \in [0, 1]$. As such, for a sufficient increase in c_P, Δ_O attenuates toward zero.

2. Differentiating Δ_B and Δ_O with respect to r yields:

$$\frac{\partial \Delta_B}{\partial r} = 0$$

$$\frac{\partial \Delta_O}{\partial r} = \frac{\eta_Q + \eta_P}{c_B c_P} \begin{cases} > 0 & \text{if } \eta_Q + \eta_P > 0 \\ < 0 & \text{if } \eta_Q + \eta_P < 0 \end{cases}$$

   This implies that if $\frac{\partial \Delta}{\partial r} > 0$ if $\Delta_O > 0$ and $\frac{\partial \Delta}{\partial r} < 0$ if $\Delta_O < 0$.

3. Differentiating Δ with respect to η_Q yields $\frac{\partial \Delta}{\partial \eta_Q} = \frac{r}{c_B c_P} > 0$.

4. Differentiating Δ with respect to η_P yields $\frac{\partial \Delta}{\partial \eta_P} = \frac{r}{c_B c_P} > 0$.■

A.1.2 Bias in Effort and Inequality in Outputs

Defining Inequality in Outputs: The model implies *inequality in outputs* as a second quantity measuring bias beyond *bias in effort*. Bias in effort is given by Proposition 1.

Inequality in outputs considers differences in expectation of the ultimate levels of service provision by group. Service provision, $\mathcal{S}$ is given by:

$$\mathcal{S}(e, q) = \begin{cases} e + (1-e)\frac{S+\gamma_P+1}{c_P} & \text{if } q = 1 \\ e + (1-e)\frac{S+\gamma_P}{c_P} & \text{if } q = 0 \end{cases}$$

I measure inequality in outputs as the difference in the expectation of service provision for each group. Define the expectation of service provision for a member of group g as:

$$\begin{aligned}\mathbb{E}[\mathcal{S}_g] =& F_g\left(\frac{b}{c_P}\right)\mathcal{S}_g(e_g^*,1)+\left(1-F_g\left(\frac{b}{c_P}\right)\right)\mathcal{S}_g(e_g^*,0)\\ =&\frac{S+\gamma_P^g}{c_P}+\frac{e_g^*(c_P-S-\gamma_P^g)}{c_P}+F_g\left(\frac{b}{c_P}\right)\frac{1-e_g^*}{c_P}\end{aligned}$$

Inequality in outputs for an individual from each group x and y is therefore defined as:

$$\begin{aligned}\mathbb{E}[\mathcal{S}_x]-\mathbb{E}[\mathcal{S}_y] =& \frac{S+\mathbb{E}[\gamma_P^x]}{c_P}+\frac{e_x^*(c_P-S-\mathbb{E}[\gamma_P^x])}{c_P}+F_x\left(\frac{b}{c_P}\right)\frac{1-e_x^*}{c_P}-\\ &\left(\frac{S+\mathbb{E}[\gamma_P^y]}{c_P}+\frac{e_y^*(c_P-S-[\gamma_P^y])}{c_P}+F_y\left(\frac{b}{c_P}\right)\frac{1-e_y^*}{c_P}\right)\\ =&\frac{-r(n_Q+2\mathbb{E}[\gamma_P^y])(2S+\eta_Q+2F_y(\frac{b}{c_P})+\eta_P+\mathbb{E}[\gamma_P^2]-c_P)}{c_P^2}+\\ &\frac{\mathbb{E}[\gamma_B^x](-F_x(\frac{b}{c_P}-\mathbb{E}[\gamma_P^x])+\mathbb{E}[\gamma_B^y](F_y(\frac{b}{c_P})+\mathbb{E}[\gamma_P^y])-S\eta_B+c_P\eta_B+c_B(\eta_Q+\eta_P)}{c_P}\end{aligned}$$

A.1.3 Endogenous Requests for Service

In the text, citizen requests of service are treated as exogenous. In this extension, I consider equilibrium levels of effort when citizens pay a cost to request a service. This adds a first step to the sequence presented in the main text, in which citizens request the service or not, denoted $R \in \{0,1\}$. Preceding the bureaucrat's decision to exert effort, this extension includes:

1. Citizen chooses whether or not to request service.

The cost of requesting a service, ξ, is proportional to c_C. The cost for requesting service is thus ξc_C.

Denote the ex-ante expected utility for citizens that would complain and citizens that would not complain, as $\mathcal{C}$ and $\mathcal{N}$:

$$\mathbb{E}[U_C^{\mathcal{C}}] = eb + (1-e)\left(b\frac{S+\gamma_P^g+1}{c_P} - c_C\right) - \xi c_C$$
$$\mathbb{E}[U_C^{\mathcal{N}}] = eb + (1-e)\left(b\frac{S+\gamma_P^g}{c_P}\right) - \xi c_C$$

Given that the citizen will only complain if $c_C < \frac{b}{c_P}$, clearly $\mathbb{E}[U_C^{\mathcal{C}}] \geq \mathbb{E}[U_C^{\mathcal{N}}]$. The citizen's expected utility is decreasing linearly in c_C. For interior e_g^*:

$$\frac{\partial\mathbb{E}[U_C^{\mathcal{C}}]}{\partial c_C} = -1 + \frac{\gamma_B^g}{c_B} + \frac{r(\gamma_P^g + S + 1)}{c_B c_P} - \xi < 0, \qquad \frac{\partial^2\mathbb{E}[U_C^{\mathcal{C}}]}{\partial c_C^2} = 0$$
$$\frac{\partial\mathbb{E}[U_C^{\mathcal{N}}]}{\partial c_C} = -1 + \frac{\gamma_B^g}{c_B} + \frac{r(\gamma_P^g + S)}{c_B c_P} - \xi < 0, \qquad \frac{\partial^2\mathbb{E}[U_C^{\mathcal{N}}]}{\partial c_C^2} = 0$$

There are three cases, that I characterize in terms of cut points in the cost of requesting service, ξ. I denote these cut points as $\bar{\xi}$ and $\underline{\xi}$ where $0 \leq \underline{\xi} \leq \bar{\xi}$. Assume, without loss of generality, that the upper support of f_g exceeds $\frac{b}{c_P}$.[1]

Case 1: $\xi \geq \bar{\xi}$: There exists some cost $\bar{\xi} \geq 0$ which makes a citizen for whom $c = \frac{b}{c_P}$ is indifferent between requesting the service and not. Substituting e_g^* and $c_C = \frac{b}{c_P}$ into $\mathbb{E}[U]_C^{\mathcal{C}} = 0$ and solving for $\bar{\xi}$ yields:

$$\bar{\xi} = S + \gamma_P^g + \frac{(c_P - S - \gamma_P^g)(r(S+\gamma_P^g+1) + c_P\gamma_B^g)}{c_B c_P}$$

Case 2: $\xi \in (\underline{\xi}, \bar{\xi})$: Denote the upper support of f_g as $\bar{c}_C$. There exists some threshold $\tilde{c}_C \in$

[1] If this assumption does not hold, the solution collapses to Case 1.

$(\frac{b}{c_P}, \bar{c}_C)$. A citizen for whom $c_C \leq \tilde{c}_C$ will request service, while a citizen for whom $c_C > \tilde{c}_C$ will not.

Case 3: $\xi \leq \underline{\xi}$: There exists some cost $\underline{\xi} \geq 0$ that makes a citizen with cost $\bar{c}_C$ indifferent between requesting the service and not. Substituting e^*_g and $c_C = \bar{c}_C$ into $\mathbb{E}[U]^{\mathcal{N}}_C$ and solving for $\underline{\xi}$ yields:

$$\underline{\xi} = \frac{r(S+\gamma_P^g) + C_P\gamma_B^g}{c_P c_B} + \frac{b(c_B c_P(S+\gamma_P^g) + (c_P - S - \gamma_P^g)(r(S+\gamma_P^g) + c_P\gamma_B^g))}{c_B c_P^2 \bar{c}_C} - 1$$

Proposition A1. *Equilibrium effort with costly requests. If, within a group g, the bureaucrat's tastes, politician's tastes, and citizen costs are independent, as the cost of requesting service, ξ, increases, the proportion of citizens requesting service decreases. The expectation of equilibrium effort exerted by the bureaucrat on each request increases (piecewise) in ξ.*

Proof:

To characterize the share of citizens in a group g that request service and the expectation of equilibrium service provided, I assume that $Cov(\gamma_B^g, \gamma_P^g) = 0$, $Cov(\gamma_B^g, C_g) = 0$, and $Cov(\gamma_P^g, C_g) = 0$ and consider each case.

Case 1: $\xi > \bar{\xi}$: For any $\xi > \bar{\xi}$, conditional on requesting service ($R = 1$), the expectation of equilibrium effort for group g is:

$$\mathbb{E}[e^*_g | R = 1] = \frac{\mathbb{E}[\gamma_B^g]}{c_B} + \frac{r(S + 1 + \mathbb{E}[\gamma_P^g])}{c_B c_P}$$

and the proportion of group g that requests the service is given by:

$$F_g\left(\frac{b(c_B c_P(S+\gamma_P^x+1) + (c_P - (S+\gamma_P^x+1))(r(S+\gamma_P^x+1) + c_P\gamma_B^x))}{c_P(-r(S+\gamma_P^x+1) - c_P\gamma_B^x + c_B c_P(1+\xi))}\right) \leq F_g\left(\frac{b}{c_P}\right)$$

Case 2: $\xi \in (\underline{\xi}, \bar{\xi}]$: For any $\xi \in (\underline{\xi}, \bar{\xi})$, conditional on requesting service ($R = 1$), the expectation of equilibrium effort for group g is:

$$\mathbb{E}[e_g^*|R = 1] = \frac{\mathbb{E}[\gamma_B^g]}{c_B} + \frac{r\left(S + \frac{F_g(\frac{b}{c_P})}{F_g(\tilde{c}_C)} + \mathbb{E}[\gamma_P^g]\right)}{c_B c_P}$$

Because $\tilde{c}_C \in (\frac{b}{c_P}, \bar{c}_C)$, $\frac{F_g(\frac{b}{c_P})}{F_g(\tilde{c}_C)} \in (F_g(\frac{b}{c_P}), 1)$. The proportion of group g that requests the service is given by $F_g(\tilde{c}_C)$.

Case 3: $\xi < \underline{\xi}$: For any $\xi \leq \underline{\xi}$, conditional on requesting service, ($R = 1$), the expectation of equilibrium effort for group g is:

$$\mathbb{E}[e_g^*|R = 1] = \frac{\mathbb{E}[\gamma_B^g]}{c_B} + \frac{r(S + F_g(\frac{b}{c_P}) + \mathbb{E}[\gamma_P^g])}{c_B c_P}$$

This is equivalent to the equilibrium with exogenous (costless) requests characterized in the main text. As was assumed in the paper, all citizens request service.

Collectively this analysis implies that the proportion of citizens requesting the service decreases in ξ and that the expectation of equilibrium effort, conditional on requesting the service, increases (piecewise) in ξ. ■

A.1.4 Bias in Effort and Inequality in Outputs

The mapping from bias in effort to inequality in outputs depends on the composition of the bias in effort (between tastes and complaint-driven bias). For this analysis, make two simplifying assumptions. First, within a group g, the distribution of tastes and costs is independent, i.e. $Cov(\gamma_B^g, \gamma_P^g) = 0, Cov(\gamma_B^g, C_g) = 0$, and $Cov(\gamma_P^g, C_g) = 0$. Second, normalize the tastes and likelihood of complaint for group y to 0, i.e. $\mathbb{E}[\gamma_P^y] = 0, \mathbb{E}[\gamma_B^y] = 0$, and $F_y(\frac{b}{c_P}) = 0$.

Bias in effort: As in Proposition in the main text, bias in effort is given by:

$$\Delta = \frac{\eta_B}{c_B} + \frac{r(\eta_P + \eta_Q)}{c_B c_P} \tag{A.7}$$

Inequality in outputs: Following the definition in Section A1.2, inequality in outputs simplifies to:

$$\mathbb{E}[\mathcal{S}_x] - \mathbb{E}[\mathcal{S}_y] = \frac{(\eta_P + \eta_Q)(rc_P - r(2S + \eta_P + \eta_Q) + c_P(c_B - \eta_B)) + \eta_B c_P(c_P - S)}{c_P^2 c_B} \tag{A.8}$$

$$= \underbrace{\frac{\eta_B}{c_B} + \frac{r(\eta_P + \eta_Q)}{c_P c_B}}_{\Delta} + \frac{(\eta_P + \eta_Q)(-r(2S + \eta_P + \eta_Q) + c_P(c_B - \eta_B)) - c_P(\eta_B S)}{c_P^2 c_B} \tag{A.9}$$

While the sign of the remaining term in the sum is ambiguous, it can be shown that within the parametric assumptions of the model, bias in effort Δ and inequality in outputs $\mathbb{E}[\mathcal{S}_x] - \mathbb{E}[\mathcal{S}_y]$ must share the same sign. Rewriting:

$$\mathbb{E}[\mathcal{S}_x] - \mathbb{E}[\mathcal{S}_y] = \frac{\eta_B(c_P - 1 - S - \eta_Q - \eta_P)}{c_B c_P} + \frac{(\eta_P + \eta_Q)(c_P(c_B + r) - r(2S - \eta_P - \eta_Q))}{c_P^2 c_B}$$

Because $c_P > S + 2$, $r \in (0, 1)$, and $c_B > 1$, the final term must maintain the same sign as $\eta_Q + \eta_P$.

Remark A1: *Given a neutral bureaucrat,* $\eta_B = 0$*, or a bureaucrat favoring the same group as that favored by oversight,* $sign(\eta_B) = sign(\eta_P + \eta_Q)$*, bias in effort favoring group* g *is sufficient to ensure that inequality in outputs favor* g*.*

A.1.5 Implications for Complaint Rates in the Observational Data

Here I derive the expected rate of complaint per the baseline model in which the request for service is exogenous. Assume that $Cov(\gamma_B^g, C_g) = 0$, and $Cov(\gamma_P^g, C_g) = 0$. The rate of complaint by group is thus given by:

$$\underbrace{F_g\left(\frac{b}{c_P}\right)}_{\text{Share "complainers"}} \underbrace{\left[1-\left(\frac{\mathbb{E}[\gamma_B^g]}{c_B}+\frac{r}{c_Bc_P}(\mathbb{E}[\gamma_P^g]+S+1)\right)\right]}_{\text{Share of "complainers" receiving service}}$$

The difference in rate of complaint, between groups x and y is thus:

$$\underbrace{\left[F_x\left(\frac{b}{c_P}\right)-F_y\left(\frac{b}{c_P}\right)\right]\left(\frac{c_Bc_P-r(S+1)}{c_bc_P}\right)}_{\geq 0} - F_x\left(\frac{b}{c_P}\right)\left(\frac{\mathbb{E}[\gamma_B^x]}{c_B}+\frac{rE[\gamma_B^x]}{c_Bc_P}\right) + F_y\left(\frac{b}{c_P}\right)\left(\frac{\mathbb{E}[\gamma_B^y]}{c_B}+\frac{rE[\gamma_B^y]}{c_Bc_P}\right)$$

In equilibrium, we should see more complaints from the group with a higher likelihood of complaint if:

$$\left[1-\frac{F_y(\frac{b}{c_P})}{F_x(\frac{b}{c_P})}\right]\left(\frac{c_Bc_P-r(S+1)}{c_bc_P}\right)+\frac{F_y(\frac{b}{c_P})}{F_x(\frac{b}{c_P})}\left(\frac{\mathbb{E}[\gamma_B^y]}{c_B}+\frac{rE[\gamma_B^y]}{c_Bc_P}\right) > \frac{\mathbb{E}[\gamma_B^x]}{c_B}+\frac{rE[\gamma_B^x]}{c_Bc_P}$$

The implication of this expression is that if the rate of observed complaint is higher for group x (with lower costs of complaint), it must be the case that the tastes of the bureaucrat and politician must not favor group x by too large of a magnitude. If this were the case, the difference in the rate at which citizens of group x receive service over those from group y would yield a higher rate of complaint from group y given the lower rate of service provision.

A.2 Cross-National Data

A.2.1 Survey Data

I refer to rates of contact with street-level bureaucrats using cross-national survey data from LAPOP AmericasBarometer (2014) and AfroBarometer (2016). The sample of countries includes all surveyed countries in Latin America, the Caribbean, and Africa with a Polity IV score above 0 in 2015. This includes both hybrid and democratic regimes.

Figure A.1 depicts the proportion of citizens reporting that they sought service from a street-level bureaucrat in the last twelve months.

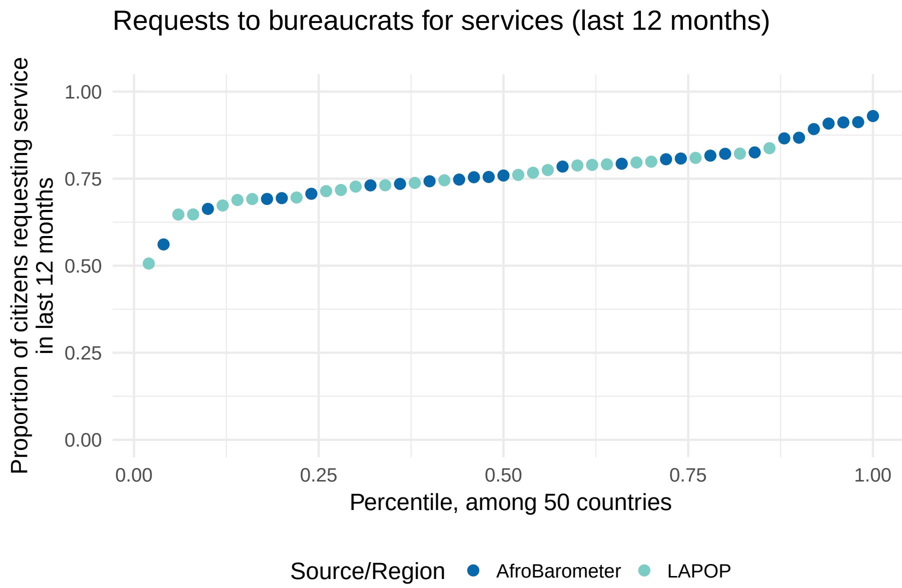


Figure A.1: The proportion of survey respondents that sought government services within the last year. Data is from AmericasBarometer (LAPOP) (2014) and AfroBarometer (2016).

A.2.1.1 AmericasBarometer (2014)

The measure of contact with street-level bureaucrats is constructed from four questions, as follows. Note that these questions are two parts; the second part of the question asks about petty corruption (bribes). If the answer indicates that the individual "had dealings" with the relevant entity, I code this as having sought a service.

1. **EXC11** "In the last twelve months, did you have any official dealings in the municipality/local government?"
2. **EXC14** "In the last twelve months, did you have any dealings with the courts?"
3. **EXC15** "Have you used any public health services in the last twelve months?"
4. **EXC16** "Have you had a child in school in the last twelve months?"

The proportion of individuals "seeking service" is the proportion of individuals that answered affirmatively to any of these questions.

The countries included in the graph are: Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Trinidad and Tobago, Uruguay, and Venezuela.

To justify the claim that interactions with street-level bureaucrats are the "most common form" of citizen-government interaction, I compare the rate of individuals seeking a service in the last twelve months (circles) to those citizens who have ever sought assistance from an elected official ($\triangle$'s) in Figure A.2. The latter are calculated from the question:

CP4a: "In order to solve your problems have you ever requested help or cooperation from a local public official or local government: for example, a mayor, municipal council, councilman, provincial official, civil governor or governor?"

Figure A.2 suggests that the proportions engaging bureaucrats in the last 12 months far exceeds the proportion that have *ever* engaged a local elected official (triangles).

A.2.1.2 AfroBarometer (2016)

The measure of contact with street-level bureaucrats is constructed from five (largely parallel) questions, as follows:

1. **Q55A** " In the past 12 months have you had contact with a public school?"
2. **Q55C** "In the past 12 months have you had contact with a public clinic or hospital?"
3. **Q55E** "In the past 12 months have you tried to get an identity document like a birth certificate,

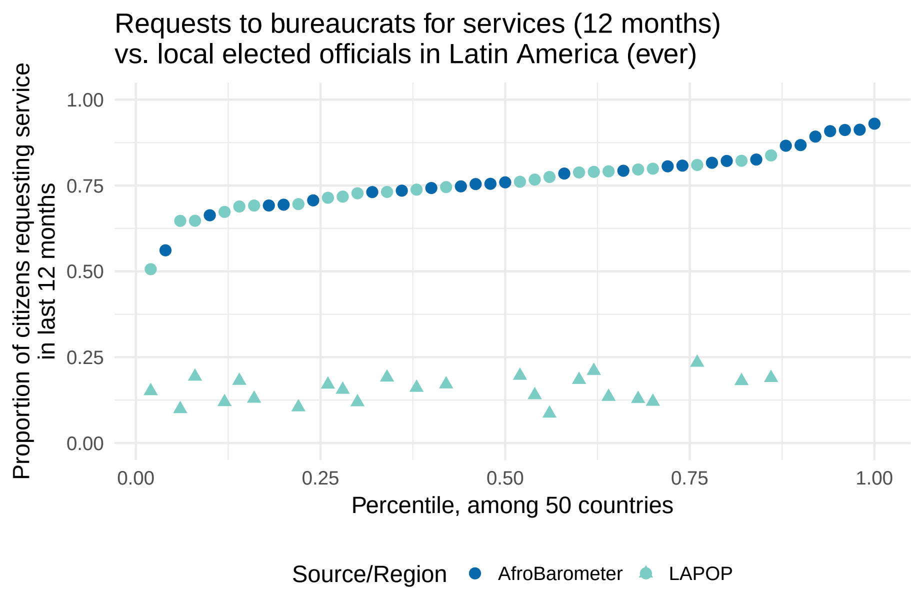


Figure A.2: The proportion of survey respondents that sought government services within the last year (circles) compared to the proportion of survey respondents from Latin America that have ever sought the services of a municipal politician.

driver's license, passport or voter's card, or a permit, from government?"

4. **Q55G** "In the past 12 months have you tried to get water, sanitation or electric services from government?"

5. **Q55K** "In the past 12 months have you had contact with the courts?"

The proportion of individuals "seeking service" is the proportion of individuals that answered affirmatively to any of these questions.

The countries included in the graph are: Algeria, Benin, Botswana, Burkina Faso, Cabo Verde, Gabon, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritius, Mozambique, Namibia, Niger, Nigeria, Senegal, Sierra Leone, South Africa, Tanzania, Tunisia, Zambia, and Zimbabwe.

AfroBarometer includes an assessment of the difficulty of obtaining service, conditional on having sought it. For these measures, I consider assessments of public administrators – those disbursing identity documents or working at the tax authority (questions Q55E, Q70A, and Q70B).[2] I average perceptions across the three questions. Where answers are missing, I average the assessments that are reported. I examine the bivariate relationship between this assessment and education graphically in Figure A.3. Citizens perceive services to be relatively difficult to access. This difficulty declines somewhat in educational attainment.

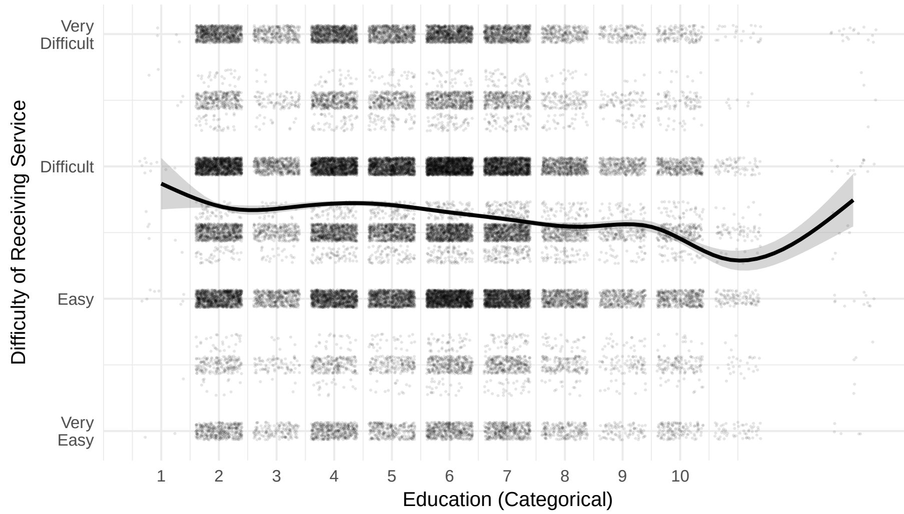


Figure A.3: Perceptions of the difficulty of obtaining services from public administrators by level of educational attainment. $n = 50,758$.

A.2.2 World Bank Worldwide Governance Indicators

The World Bank Worldwide Governance Indicators (WGI) allows for cross-national characterizations of governance, broadly conceived. Some of these measures measure bureaucratic outputs.

[2]The tax questions are as follows. **Q70A**:“Based on your experience, how easy or difficult is it to obtain the following services from government? Or do you never try and get these services from government: To find out what taxes and fees you are supposed to pay to the government?” **Q70B**:“Based on your experience, how easy or difficult is it to obtain the following services from government? Or do you never try and get these services from government: To avoid paying the income or property taxes that you owe to government?”

The intent in providing this information is simply to demonstrate where Colombia ranks globally in standard measures of governance. Using data from 2016 (the most recent WGI data) evaluate Colombia's rank, as a percentile, among:

- All countries in the World Bank WGI data ($n = 214$).
- All Spanish, French, and Portuguese-speaking countries in Latin America ($n = 19$).
- All OECD countries ($n = 37$). Note that Colombia joined the OECD in July 2018.

Table A.1 shows Colombia's rank among the three comparison groups for each of three indicators. While Colombia unsurprisingly performs quite poorly relative to all reference groups on the "Political Stability and Absence of Violence/Terrorism" measure, the other indicators which are plausibly more relevant measures of bureaucratic outputs. In general, Colombia generally performs somewhere around the median of all countries, in the top tercile of Latin American countries, and the bottom decile of OECD countries.

	Colombia's percentile among . . .		
	World	Latin America	OECD
Indicator	($n = 214$)	($n = 19$)	($n = 37$)
Political Stability and Absence of Violence/Terrorism	13.81	5.56	2.78
Rule of Law	41.35	72.22	5.56
Control of Corruption	44.23	72.22	5.56
Voice and Accountability	49.75	55.56	5.56
Government Effectiveness	54.33	66.67	2.78
Regulatory Quality	67.31	77.78	11.11

Table A.1: Colombia's rankings on each World Bank Worldwide Governance Indicator (2016) as a percentile within the relevant comparison group.

A.3 Original Survey of Bureaucrats

The manuscript cites one descriptive finding from an original survey of street-level bureaucrats in *alcaldiías* in Bogotá and Cundinamarca. The details of the survey are as reported here. The survey was conducted in October and November, 2016. Two parallel surveys were conducted: one of citizens awaiting service and one of bureaucrats providing service in select entities. The surveys were conducted in:

- *Alcaldías*: local *alcaldías* in Bogotá and municipal *alcaldías* in Cundinamarca
- CADES/SUPERCADES: These are District (Bogotá) Centers for Public Service where citizens can seek many public services.
- Local offices of the Registraduría Nacional del Estado Civil
- Local notaries (public/private).

The relevant sample cited in the paper includes 73 surveys of bureaucrats from 14 *alcaldías*. The *alcaldías* were purposefully sampled but the timing of the visit was randomly assigned. The sampling included the first 5-8 bureaucrats that we encountered starting at reception (e.g. street level bureaucrats) that were willing to take the survey. In this sense, the sample is not random, but encompasses street-level bureaucrats in these entities present at the time of the survey.

The relevant question cited in the descriptive statistic in the paper was an open response question, enumerated and translated as follows:

- "¿Si usted tomara una decisión que su supervisor no apoyara, cuál sería la consecuencia?"
- Translation: "If you made a decision that your supervisor did not support, what would be the consequence?"

The responses ranged from verbal admonishment to more formal admonishments (in the form of a memorandum) to unwillingness to renew a contract (contractors only).

A.4 Bogotá Complaint Data

Data on formal complaints from Bogotá is collected by the Veeduría Distrital, an oversight organ of the city government. Data is available at `tablerocontrolciudadano.veeduriadistrital.`

`gov.co:3838/BogotaDashboard/`. The data consist of 464,387 PQRS petitions submitted to city entities in Bogotá between January 1, 2017 and June 30, 2018. PQRS stands for "*peticiones, quejas, reclamos, y sugerencias*," translated "petitions, complaints, [another word for] complaints, and suggestions)." These comprise represent formal written requests, not verbal or informal complaints. Note that the per capita rate of PQRS submission during this period is 5.68%, or one submission per ≈17.5 people.[3]

The PQRS are characterized by type, as in Table A.2. Note that there are more words for complaints in Spanish than in English. I focus on the first three categories (the complaints) in the subsequent analysis $n = 440,803$.

PQRS Type	Translation	n	Proportion
Denuncia	Report (of complaint)	2,501	0.005
Queja	Complaint	99,302	0.214
Reclamo	Complaint	339,000	0.730
Sugerencia o Felictación	Suggestion or congratulation	23,584	0.051

Table A.2: PQRS submitted in Bogotá, January 2017-June 2018. The type designation is made by the Veeduria (or receiving entities). Translations by author.

Of the complaints, 63,330 were registered by *alcaldías locales*, the entities audited in the experiment. Other complaints were directed to district-wide entities. To assess the correlation between class and propensity to complain, I examine the relationship between the relative wealth of a locality and the per-capita rate of complaint submission. To measure the wealth of a locality, I examine the average *estrato* (strata) of all residential properties. Strata range from 1 (very low/*bajo bajo*) to 6 (very upper/*alta alta*). While these zoning designations are technically made to properties, citizens identify *estrato* with class. Equating the two implies an assumption that lower-class Bogotanos are priced out of rich neighborhoods/dwellings and few middle- and upper-class Bogotanos choose to

[3]Calculated based on an estimated population of 8,181,047 residents.

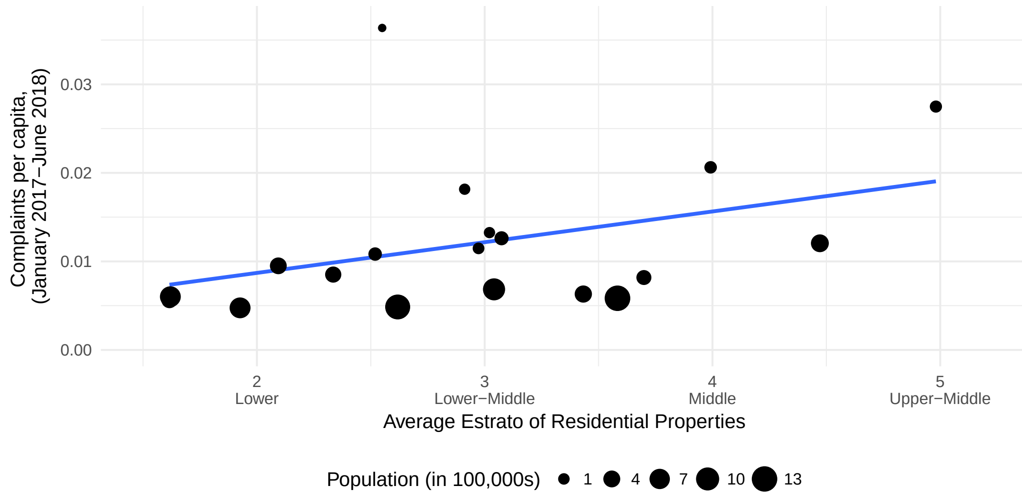


Figure A.4: Rate of complaints filed by locality in Bogotá by average wealth of the locality.

live substantially below their means. The marked degree of differentiation of localities in average *estrato* implies high levels of residential segregation.

Figure A.4 plots the rate of complaint submission by wealth of localities. There is a clear positive relationship between the wealth of the locality and the rate of complaint submission. This occurs despite the fact that service is believed to be better in wealthy localities, suggesting that this analysis *understates* the relationship between class and propensity to complain. The outlier, La Candelaria, is a very small locality in the center of Bogotá with a vastly disproportionate tourist/foreigner ("ex-pat") population. To the extent that "ex-pats" (foreigners from rich countries) choose to live in a locality with a relatively low *estrato*, resident wealth is understated by the *estrato* designation in La Candelaria.

A.5 Demographic Data Related to Bias Treatments

Colombia's last national census was conducted in 2005. A new census will be conducted this year (2018). Given the vintage of the data, I use population projections where relevant and available.

Otherwise, this section describes historical trends through 2005. All aggregate data in this section comes from the Departamento Administrativo Nacional de Estadística and all microdata comes from IPUMs.

A.5.1 Regional Accents

In this section, I describe the geographic coverage and prevalence of the three accents utilized in the audit experiment. Figure A.5 shows the departments to which these (generalized) accents are native. These are among the most densely populated regions of Colombia.

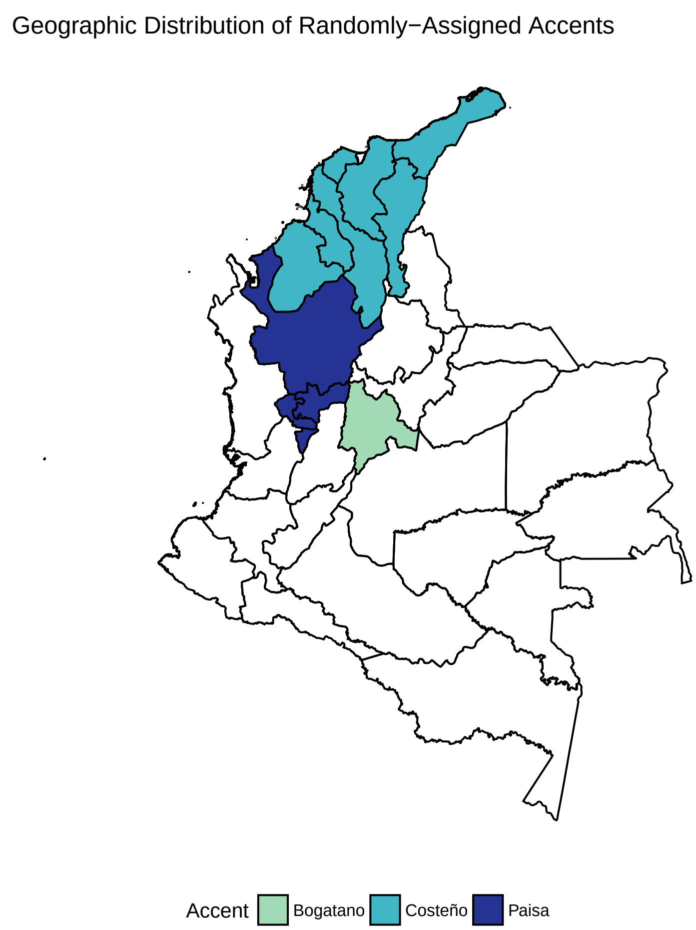


Figure A.5: Geographical coverage of the Bogotano, Costeño, and Paisa accents at the departmental level. Note that the map of the Bogotá accent includes the department of Cundinamarca. Some portion of Cundinamarca's population speaks with a different accent (Cundinaboyacense).

I approximate the number of speakers of each accent in Colombia in Figure A.6.[4] To approximate these quantities, I consider the most widely-spoken accent in each department. I aggregate the projected population (2017) by DANE for each department and sum across the departments associated with each accent. This exercise indicates that nearly 60% of the Colombian population speaks one of the three regional accents utilized in the audit experiment.

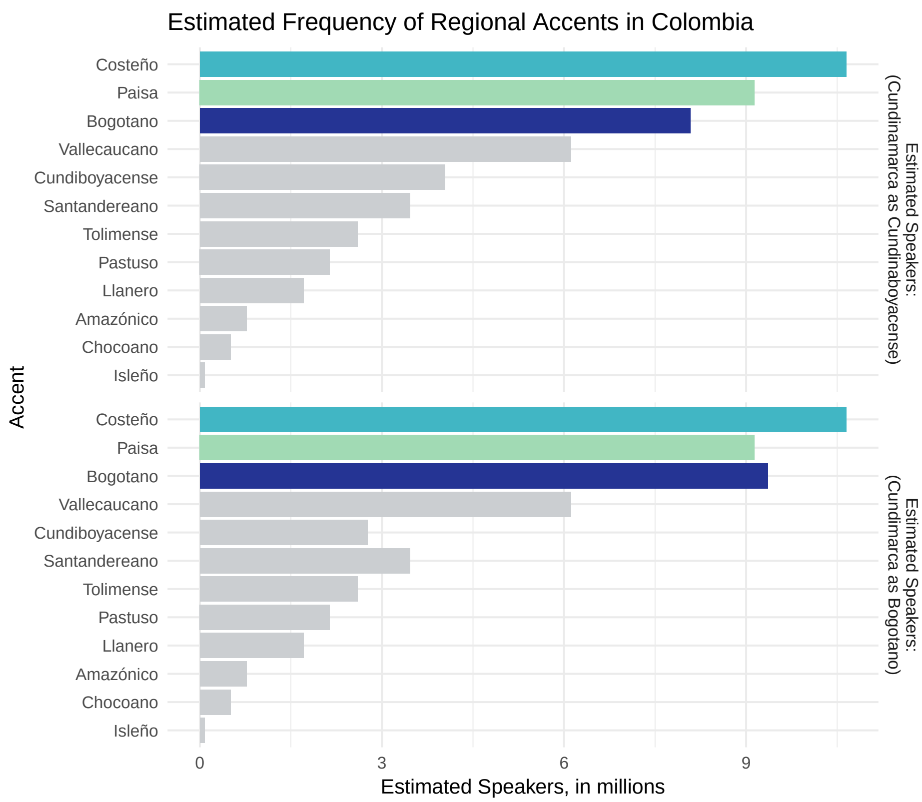


Figure A.6: Approximate number of speakers of Colombia's twelve major accents. The graph shows that the most widely-spoken accents are the Bogotano, Costeño, and Paisa accents used in the experiment. Approximately 60% of the Colombian population speaks one of these three accents. The panels differ in the classification of the accent in the department of Cundinamarca (Bogotano or Cundinaboyacense).

[4]Note that several of the accents are also spoken in neighboring countries, e.g. the Llanero accent in Venezuela. These counts only include speakers of the accent in Colombia.

A.5.2 Socioeconomic Class

Detailed data on socioeconomic class is not available in Colombia. As such, I present the distribution of Colombians by class as per the 2005 census in Figure A.7.

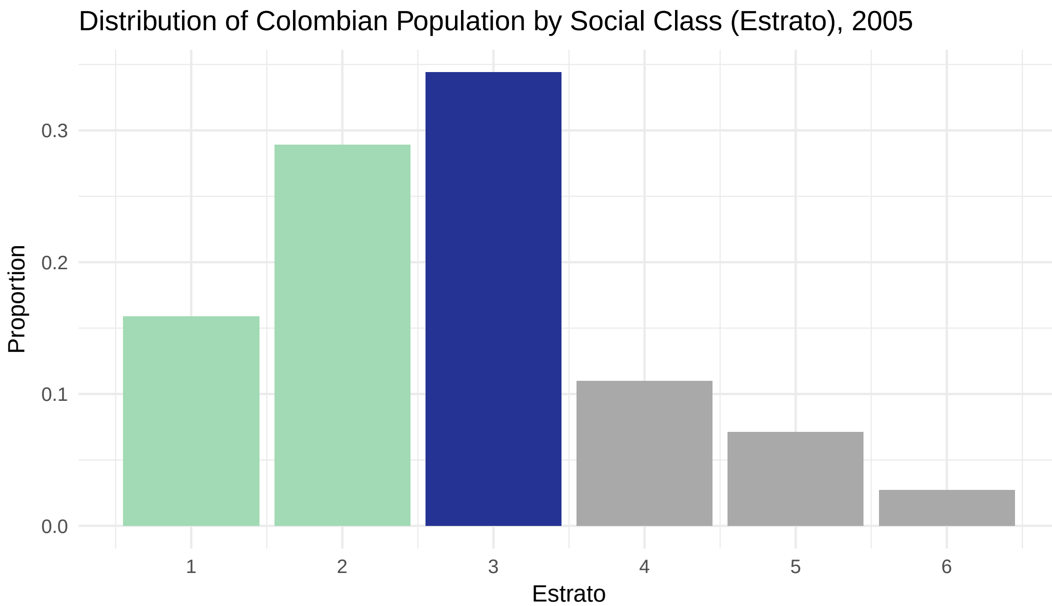


Figure A.7: Distribution of classes in the 2005 Colombian census. The social class of the treatments are denoted by the colored bars.

A.5.3 Migration

In Table A.3, I place lower bounds on the proportion of Colombians that have (a) ever migrated and (b) migrated within the last five years using census microdata subsamples available from IPUMs. I bound the share of Colombians that have ever migrated by examining the share of individuals residing in their municipality of birth. Note that this is a lower bound as reverse-migrants will not be counted as migrants. I bound the share of individuals who report migrating to a different geographical unit in the past five years. This similarly does not count reverse-migrants and should be considered a lower bound.

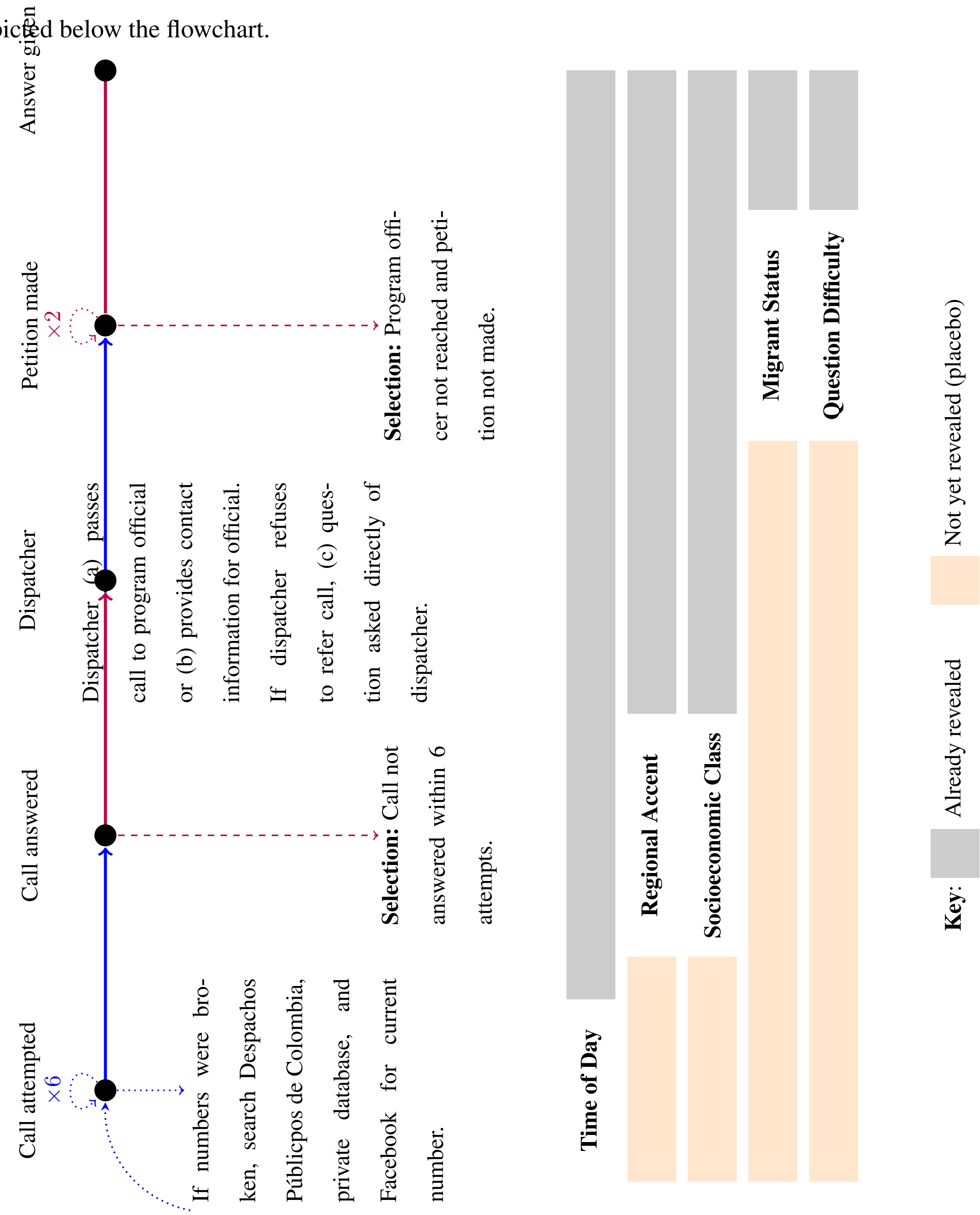
In general, the bounds suggest that at least 37% of the Colombian population (2005) has migrated at some point in their lives. This remains relatively stable across time. Furthermore, in past censuses, 10-17% of Colombians report migrating within the past five years. Note that migration to other departments is slightly more common than migration within the same department, though both forms of migration are common. These totals include both ordinary and conflict-induced migration.

			Census of		
	1964	1973	1985	1993	2005
A: MICRODATA SAMPLE SIZE (FROM IPUMS)					
Sample Size	349,652	1,988,831	2,643,125	3,213,657	4,006,168
B: LIFETIME MIGRATION (LOWER BOUND), RESIDES IN:					
Municipality of Birth	0.635	0.626	0.642	0.594	0.623
Same Department, Different Municipality	0.180	0.160	0.145	0.170	0.152
Different Department	0.181	0.209	0.210	0.225	0.204
(Born Abroad)	0.004	0.004	0.003	0.003	0.003
No Response	0.000	0.000	0.000	0.007	0.019
C: MIGRATION IN LAST FIVE YEARS (LOWER BOUND):					
Has not Migrated			0.869	0.834	0.905
Migrated within Department			0.056	0.062	0.031
Migrated from a Different Department			0.069	0.084	0.041
Immigrated			0.006	0.004	0.002
No response			0.000	0.016	0.021

Table A.3: Lower bound on lifetime migration and migration within the last five years in Colombian censuses since 1964. Migration within the last five years was not included in the 1964 or 1973 censuses. Population-weighted means estimated from census microdata.

A.6 Call Sequencing Flow Chart

Confederates were trained, in part, by memorizing a basic flowchart for each call which mirrored the instrument that they filled in. This graphic provides a translated and vastly simplified version of the sequence of calls, denoting the point in the call at which each factor was revealed and delineating the subsamples. As in the main manuscript three samples are defined and temporally delineated, as depicted below the flowchart.



A.7 Petitions, Correct Answers

The following table lists the questions and correct answers used in the audit experiment translated to English. There are eight total questions. Because the difficulty of the question and migrant status are conveyed in the petition, there is effectively a 2×2 for each of the two programs, SISBÉN and MFA. The petitions appear in Tables A.4 and A.5, respectively.

All questions were presented in the third person to minimize detection (e.g. a request for the petitioner's identification number). Empirically, a substantial proportion of observed calls in government call centers were made on behalf of someone else. Further, in piloting, the responses to first person versus third person requests were qualitatively similar with the exception of petitions for an identification number.

	Program	Difficulty	Migrant	Question	Correct Answer
1	SISBÉN	Easy	Migrant	I am doing my neighbor a favor and asking about SISBÉN. She just moved to this municipality and wants to register for SISBÉN. What does she need to do to register?	1. Go to the SISBÉN office. 2. Ask for an application for the survey for the first time. 3. She must be a resident of the home, older than 18 years, and present her identification card. 4. (At this time, we are not doing new registrations for SISBÉN.)
2	SISBÉN	Easy	–	I am doing my neighbor a favor and asking about SISBÉN. She wants to register for SISBÉN. What does she need to do to register?	1. Go to the SISBÉN office. 2. Ask for an application for the survey for the first time. 3. She must be a resident of the home, older than 18 years, and present her identification card. 4. (At this time, we are not doing new registrations for SISBÉN.)
3	SISBÉN	Difficult	Migrant	I am doing my neighbor a favor and asking about SISBÉN. She just moved to the municpality and tried to enter *Colombia Mayor* and they did not let her. She has a score of 45. What can she do to lower her score? She is 65 years old.	1. Go to the SISBÉN office. 2. Check the form of the person to verify that the data in the system are correct. 3. If there are differences on the form, complete a petition of disagreement. 4. For the request to be approved, she will need to ask for a new survey. 5. She must be registered for SISBÉN in this municipality and her score must correspond to the guidelines in this [category of] municipality.
4	SISBÉN	Difficult	–	I am doing my neighbor a favor and asking about SISBÉN. She tried to enter *Colombia Mayor* and they did not let her. She has a score of 45. What can she do to lower her score? She is 65 years old.	1. Go to the SISBÉN office. 2. Check the form of the person to verify that the data in the system are correct. 3. If there are differences on the form, complete a petition of disagreement. 4. For the request to be approved, she will need to ask for a new survey.

Table A.4: List of SISBÉN petitions, translated to English.

Program	Difficulty	Migrant	Question	Correct Answer
MFA	Easy	Migrant	I am doing my sister a favor and asking about MFA. She just moved to this municipality and wants to register for MFA. She has a 6 year old son and a 10 year old daughter. What does she need to do to register?	1. She must come to the office of the municipal *enlace* or point of service. 2. She must bring her document of identification. 3. She must bring the civil registration of all children under 7 and the identification card for all children between ages 7 and 18. 4. For children under 6, she should bring a certificate of children's attendence at (medical) exams of growth and development, certified with the name of the attending official. 5. For children in school, the mother should bring proof of enrollment in school. 6. (At this time, we are not doing new registrations for MFA). 7. The mother must register for SISBÉN in this municipality and have a qualifying score for MFA.
MFA	Easy	–	I am doing my sister a favor and asking about MFA. She wants to register for MFA. She has a 6 year old son and a 10 year old daughter. What does she need to do to register?	1. She must come to the office of the municipal *enlace* or point of service. 2. She must bring her document of identification. 3. She must bring the civil registration of all children under 7 and the identification card for all children between ages 7 and 18. 4. For children under 6, she should bring a certificate of children's attendence at (medical) exams of growth and development, certified with the name of the attending official. 5. For children in school, the mother should bring proof of enrollment in school. 6. (At this time, we are not doing new registrations for MFA). 7. The mother must register for SISBÉN in this municipality and have a qualifying score for MFA.
MFA	Difficult	Migrant	I am doing my neighbor a favor and asking about MFA. She just moved to this municipality. How does she change her registration with the program?	1. Go to the MFA office. 2. She must present a signed written request documenting that she has moved to the municipality. 3. The mother must register for SISBÉN in this municipality and have a qualifying score for MFA.
MFA	Difficult	–	I am doing my neighbor a favor and asking about MFA. Her sister, who was a MFA recipient died and left her in charge of her nephew. How does she become the MFA recipient for her nephew?	1. Go to the MFA office. 2. She must present her identification document. 3. She must turn in the document of custody and personal care of the child, issued by the competent authority: the defender or commisary of the family. 4. She must present the document from the civil registry documenting her sister's death. 5. The mother must be registered for SISBÉN in this municipality and have a qualifying score for MFA.

Table A.5: List of MFA petitions, translated to English.

A.8 Municipal Sampling

I use stratified random sampling in an effort to maximize within variation while limiting the probability of detection. Table 2 defines the characteristics of the strata. Here, I provide additional details on sampling of municipalities. The sampling probability for medium and large municipalities is clearly 1. However, the sampling probability for small municipalities is heterogeneous since sampling was proportional to estimated population (2018). While Eframidis and Spirakis (2006) provide a numerical solution for sampling probabilities for weighted sampling without replacement, the closed form probabilities require construction of all combinations, this solution is computationally infeasible in the present application with $\binom{898}{400}$ possible samples.

I approximate the sampling probabilities for the small stratum as follows:

1. Sample 400 municipalities with weights proportional to estimated population 100,000 times.
2. Calculate the probability that each municipality was sampled across the 100,000 draws.
3. Fit the a function by regressing these probabilities on 2018 population and polynomial terms (up to a tertic term) in OLS.
4. Predict the sampling probabilities from the fitted model.

Figure A.8 depicts the sampling probabilities for municipalities as a function of population size. It demarcates the three strata by color. Note that Bogotá, the largest municipality, is divided into municipalities for the purposes of the experiment.

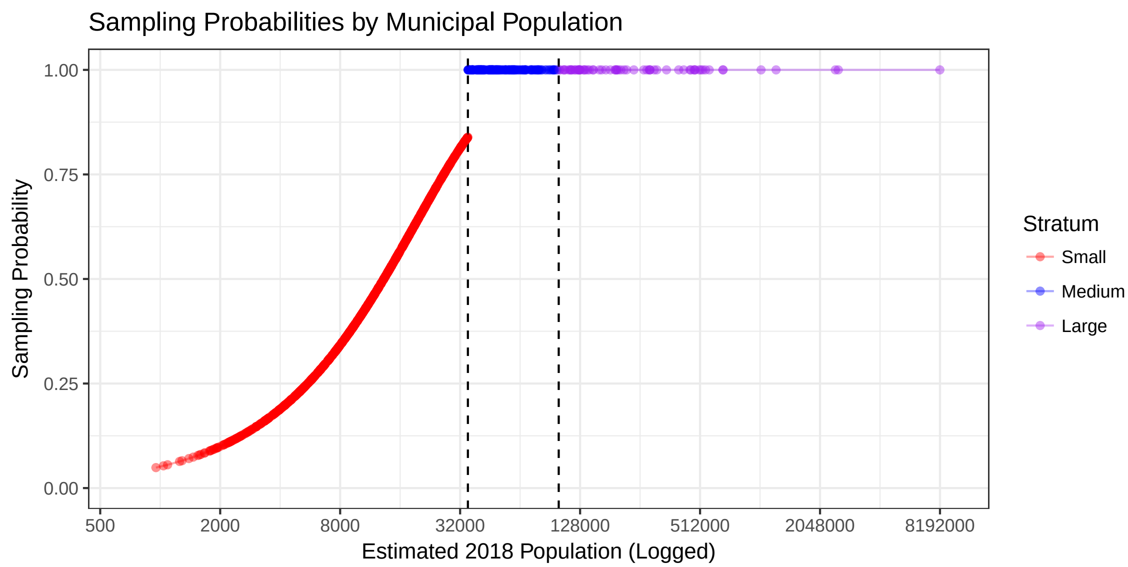


Figure A.8: The sampling probabilities for municipalities as a function of population. The points represent individual municipalities in the universe. Note that these points are municipalities; localities in Bogotá are not represented in the present graph.

Predictably, this sampling procedure gives rise to a sample that is larger, on average, than the pool of municipalities as a whole, but one that provides support across the distribution of municipal populations, as depicted in the density plots in Figure A.9. Note that in the experiment, 16 of Bogotá's localities are sampled according to the same rules for a total of 618 entities. Bogotá is represented as a whole in Figure A.9.

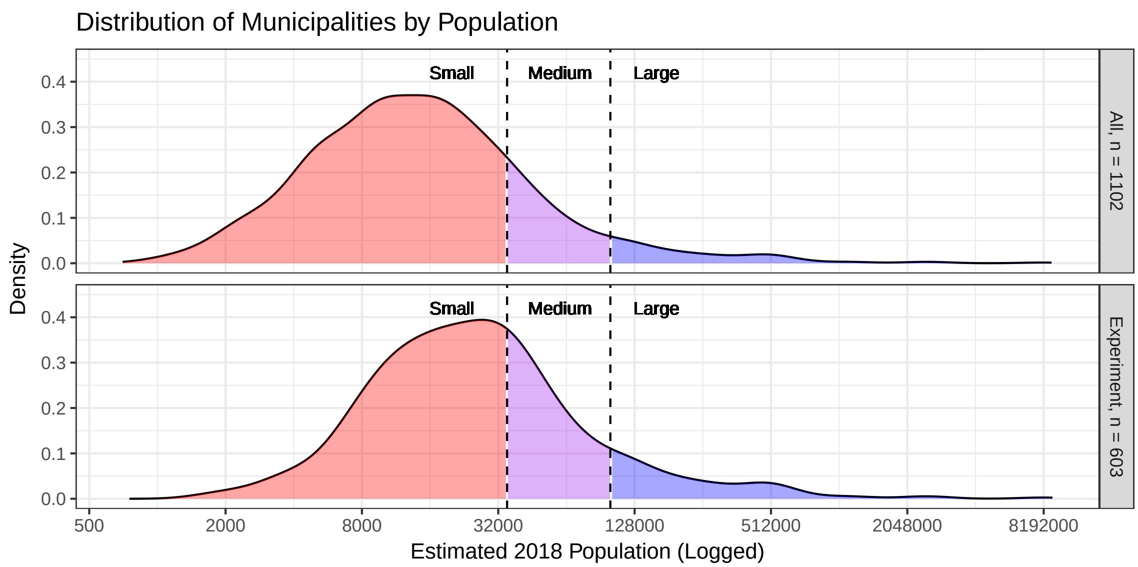


Figure A.9: The municipalities in the sample compared by population compared to the full distribution of Colombian municipalities.

A.9 Random Assignment

A.9.1 Factorial Design

The random assignment is generally blocked at the entity level (including both petitions) in order to maximize the within variation within each entity. The specifications by municipality size stratum are indicated in Table A.6.

A.9.2 Random Assignment of Enumerators

Enumerators come from each of the three regions corresponding to the regional accents. All enumerators are actresses and voice both the low- and neutral-class scripts. Within an accent, calls were block random-assigned to enumerators. As such, all enumerators cover all of the other factors of randomization.

Stratum	Attribute	Details	Restrictions
Large	Class	• 3 neutral • 3 low	Blocked by municpality/locality with each of 6 accent+ class combinations represented
	Accent	• 2 Bogotá • 2 Paisa • 2 Costeño	Blocked by municpality/locality with each of 6 accent+ class combinations represented
	Difficulty	• 4 easy • 2 hard	Blocked by program (SISBEN and MFA): easy questions (all) with one SISBEN and one MFA hard question.
	Migrant Status	• 3 migrant • 3 non-migrant	Both easy migrant questions; one hard question includes migrant line.
	Time of day	• 2-4 morning • 2-4 afternoon	Blocked by program (SISBEN, MFA) and municipality/locality
Medium	Class	• 2 neutral • 2 low	Blocked by program (SISBEN, MFA) and municipality/locality
	Accent	• 1 Bogotá • 1 Paisa • 1 Costeño • (1 repeated)	Blocked by municipality/locality. Repeated accent appears in both programs
	Difficulty	• 2 easy • 2 hard	Blocked by program (SISBEN/MFA) and municipality/locality
	Migrant Status	• 2 migrant • 2 non-migrant	Blocked by program (SISBEN/MFA) and municpality/locality
	Time	• 2 morning • 2 afternoon	Blocked by program (SISBEN/MFA) and municipality/locality
Small	Class	• 1 neutral • 1 low	Blocked by municpality/locality.
	Accent	• 2 distinct accents	Two distinct accents assigned to each municipality/locality
	Difficulty	• 1 easy • 1 hard	Blocked municipality/locality.
	Migrant Status	• 1 migrant • 1 non-migrant	Blocked by municipality/locality
	Time of day	• 1 morning • 1 afternoon	Blocked by municipality/locality

Table A.6: Factorial design by stratum with restrictions on the randomization intended to maximize within-municipality variation. Note that the total number of petitions reflects those sent to both SISBÉN and MFA (combined).

A.9.3 Random Assignment of Order of Calls

Calls were randomly assigned to an order for the original call. The assignment of the order proceeds as follows:

- Assignment to portions of the order distribution:
  - Within the large stratum (denoted $\mathcal{L}$), block randomly assign calls within each entity to six blocks. These correspond to sextiles of the rollout. Denote these blocks as $b_i^L \in \{1, 2, 3, 4, 5, 6\}$
  - Within the medium stratum (denoted $\mathcal{M}$), block randomly assign calls within each entity to four blocks. These correspond to quartiles of the rollout. Denote these blocks as $b_i^M \in \{1, 2, 3, 4\}$.
  - Within the small stratum (denoted $\mathcal{S}$), block randomly assign calls within entity to two blocks. These correspond to halves of the rollout. Denote these blocks as $b_i^S \in \{1, 2\}$
- Within each block in the rollout (defined above), randomly assign an integer ordering to the calls, denoted $O_i \in \{1, ..., \frac{|\mathcal{L}|}{6}\}$ for the large stratum, where $|\mathcal{L}|$ is the cardinality of the set of calls in the large stratum.
- Use the following formula to convert the rollout to a continuous measure between 0 and 1, shuffling the calls from the strata:

$$R_i = \frac{\mathbb{I}_{i \in \mathcal{L}}[(\frac{|\mathcal{L}|}{6}(b_i^L - 1) + O_i]}{|\mathcal{L}|} + \frac{\mathbb{I}_{i \in \mathcal{M}}[(\frac{|\mathcal{M}|}{4})(b_i^M - 1) + O_i]}{|\mathcal{M}|} + \frac{\mathbb{I}_{i \in \mathcal{S}}[\frac{|\mathcal{S}|}{2}(b_i^S - 1) + O_i]}{|\mathcal{S}|} \tag{A.10}$$

- Assign calls to enumerators (assigned as above) based on their order in the distribution.

A.10 Distribution of Treatments

The frequency with which treatment cells were utilized in the experiment is reported in Table A.7. Note that There was a higher probability of assignment to easy than hard questions in the large stratum, and thus in the experiment as a whole. Otherwise, all cells (within easy and hard) had equal probabilities of assignment.

Easy Petition								
			Migrant Accent			Non-Migrant Accent		
			Bogotano	Costeño	Paisa	Bogotano	Costeño	Paisa
Morning	Class	Lower	44	37	47	46	44	33
		Middle	57	37	50	45	48	47
Afternoon	Class	Lower	28	39	40	42	26	37
		Middle	50	52	49	32	36	32
Difficult Petition								
			Migrant Accent			Non-Migrant Accent		
			Bogotano	Costeño	Paisa	Bogotano	Costeño	Paisa
Morning	Class	Lower	25	34	28	44	32	43
		Middle	36	36	26	27	37	24
Afternoon	Class	Lower	43	38	35	41	42	40
		Middle	29	33	35	34	39	37

Table A.7: Distribution of the frequency of treatment cells in the $2 \times 3 \times 2 \times 2 \times 2$ factorial design.

A.11 Experimental Design Validation

A.11.1 Compliance

With a relatively complex audit experiment and a large team of confederates, compliance with treatment assignment is a concern for the analysis and interpretation of findings. To address such concerns, all calls were recorded. Subsequent to the experiment, two trained research assistants listened to all the calls (a full time job for over a month) and marked what they heard in the calls.

The research assistants were not apprised of the schedule of treatment assignment.[5] I examine compliance factor by factor in Table A.8.

Factor	Validation Data	Assignment	Coding in Validation data				% Compliers
Time	Phone log times				Mix with		
			Morning	Afternoon	$\geq$ 1 intent		
		Morning	98.7%	0	1.3%		98.7%
		Afternoon	0.3%	98.0%	1.65%		98.0%
Accent	Double entry		Bogotá	Costeño	Paisa	Indeterminate	
		Bogotá	99.7%	0%	0%	0.3%	99.7%
		Costeño	0.3%	98.7%	0%	1.0%	98.7%
		Paisa	0.8%	0.5%	98.4%	0.3%	98.4%
Class	Double entry		Lower	Neutral	Indeterminate		
		Lower	76.7%	11.2%	13.1%		76.7%
		Neutral	6.7%	79.3%	14.0%		79.3%
Difficulty	Double entry		Technical	Easy			
		Technical	99.3%	0.7%			99.3%
		Easy	0.8%	99.2%			99.2%
Migrant	Double entry		Migrant	Resident			
		Migrant	97.3%	2.7%			97.3%
		Resident	5.0%	95.0%			95%

Table A.8: Rates of compliance by treatment. Double entry refers to the hand coded data by outsiders listening to recordings after the fact. The phone log times were automatically recorded and outside the purview of confederates.

A.11.2 Detection

Of the 1194 answered calls, bureaucrats appeared to detect six of the calls as audits, per the classification of double coders. These calls are detailed below. Note that all calls that were detected were detected prior to the statement of the petition. There are no systematic differences in detection by municipal population stratum (as defined in the sampling of municipalities) or destination of the

[5]Calls for 10 petitions were lost by the software doing the recordings. These calls represent less than 1% of the total calls and I have no reason to believe that the missingness is systematic.

calls (department). Further, there are no systematic differences in the characteristics of the petition or petitioner except that these calls occurred later in the sequence within a given *alcaldía*.

	Stratum	Department	Call Order	Program	Time	Accent	Class	Difficulty	Migrant
1	Large	Bogotá	4/6	MFA	PM	Costeño	Lower	Easy	Resident
2	Large	Bogotá	5/6	MFA	AM	Paisa	Lower	Easy	Migrant
3	Large	Cundinamarca	5/6	MFA	PM	Paisa	Lower-Middle	Technical	Migrant
4	Medium	Bolívar	4/4	SISBÉN	PM	Paisa	Lower-Middle	Technical	Resident
5	Small	Bolívar	2/2	MFA	PM	Costeño	Lower	Technical	Resident
6	Small	Cauca	2/2	SISBÉN	PM	Paisa	Lower-Middle	Technical	Resident

Table A.9: This table documents the calls that were detected, as denoted by the double coders. Both calls detected in Bogotá were detected in the same locality (*alcaldía local*).

A.11.3 Joint Test of Interactions between Identity Treatments

The empirical strategy employed in this paper analyzes "along the margins" of the factorial experimental design. In this section, I allay concerns of substitutability or complementarities between the identity-based attributes used to measure bias. In this analysis, I utilize F-tests of the joint significance of the relevant interactions. To conduct this analysis, I use the IPW model from the main analysis as the restricted model:

$$Y_{ipm} = \sum_{j \in \mathbf{Z}} \beta_j Z_i^j + \kappa_p + \epsilon_{ipm} \tag{A.11}$$

I then specify an unrestricted model, as in Equation. Note that M_i indicates lower-middle class, R_i indicates a resident, B_i indicates a Bogotá accent, and C_i indicates a Costeño accent.

$$\begin{aligned} Y_{ipm} = \sum_{j \in \mathbf{Z}} \beta_j Z_i^j + \kappa_p + \chi_1 M_i R_i + \chi_2 M_i B_i + \chi_3 M_i C_i + \chi_4 R_i B_i + \chi_5 R_i C_i + \\ \chi_6 M_i R_i B_i + \chi_7 M_i R_i C_i + \epsilon_{ipm} \end{aligned} \tag{A.12}$$

I test the null hypothesis that $\chi_1 = \chi_2 = \chi_3 = \chi_4 = \chi_5 = \chi_6 = \chi_7 = 0$ using an F-test. The results of this test are reported in Table A.10. I fail to reject the null hypothesis for 9/10 outcomes at the $\alpha = 0.1$ level, providing no evidence that of interactions between the identity-based treatments.

Outcome	N Unrestricted	N Restricted	F-statistic	p-value
Dispatcher Gave Name	1186	1179	0.17	0.99
Petition Made	1186	1179	0.43	0.88
Second Official	1186	1179	1.24	0.28
Program Officer	1186	1179	0.23	0.98
Complete	1186	1179	1.44	0.19
Incomplete	1186	1179	0.90	0.50
Any Info.	1186	1179	1.04	0.40
No Info.	1186	1179	1.24	0.28
Alcaldía Only	1186	1179	2.71	0.01
Red Tape	1186	1179	0.80	0.58

Table A.10: Results of F-tests of the significance of interactions between identity-based characteristics for each of the main outcomes reported in Table 4.

A.12 Personnel Data

A.12.1 Source

The administrative data comes from the compilation of five administrative datasets, obtained by various means. Table A.11 documents the datasets and how they were obtained.

At the aggregate level, this data overlaps with the experimental sample of 600 *alcaldías* and 18 *alcaldías locales* as follows. “No contractor data coverage” indicates that there are no current contracts (as of the experiment) in the system.

	Contractor Data Coverage (SECOP-I, SECOP-II) Coverage	No Coverage
Civil Servant Coverage	550	22
No Civil Servant Coverage	40	6

A.12.2 Identification of Relevant Personnel

The individual data comes from two sources. First, I have a list of program officers furnished by PS (MFA) and downloaded from DNP (SISBÉN). Second, the list of names comes from the double en-

Dataset	Population	Coverage (Experimental)	Method Obtained
SIGEP	Civil Servants	>800 municipalities	Obtained from Departamento Administrativo de la Función Pública by *derecho de petición* (≈ FOIA request). Request granted in April 2018.
SIGEP	Civil Servants	798 employees in experimental *alcaldías*	Identified by hand search and scraped from online database, May 2018.
SIDEAP	Civil Servants	20 Bogotá *alcaldías locales*	Downloaded in April 2018.
SECOP-I	Contractors	1100 *alcaldías*, Metro SALUD Medellín	Downloaded in April 2018.
SECOP-II	Contractors	20 Bogotá *alcaldías*	Downloaded in April 2018.
SIDEAP	Contractors	Bogotá *alcaldías* (for cross-referencing)	Downloaded in April 2018.

Table A.11: Administrative datasets, coverage, and method used to obtain the data.

tered calls. Enumerators wrote the names shared by officials. Most names in the administrative data contain four names, two first names and two last names, as is standard in administrative documents.

The names from the calls are often much shorter (often one first name and one last name) and spelling is approximated by the research assistants. While most Spanish names are quite straightforward to spell, some Colombians have adopted English names. Spelling of these names varies substantially. For example, in my data the common name pronounced “Jason” (to an English speaker) is spelled “Jeison,” “Jeisson,” “Yeison,” and “Yeisson.” This poses a challenge for string matching. To maximize information before matching to the administrative datasets, I matched the calls to the administrative data, which would match the hypothetical the name “Yohana Díaz” to “Johanna Luz Díaz Guerrero.”

I then took the list of unique names (some full and some short) to the dataset of civil servants and contractors. To identify relevant personnel, I implemented a q-gram matching algorithm, selecting matches that maximize the number of common 3-grams within entity. I then revised the matches with the most q-grams in common by hand, applying a 90% (of the shorter name) threshold or a phonetic match (i.e. the multiple spellings of "Jason") to identify employees.

Where no match was located, I searched the online dataset of SIGEP by hand, since the coverage of the original civil servant dataset was much more limited. If the employee was located on SIGEP, I wrote a Python program to scrape their information. I applied the same rule for identification to the SIGEP searches as to the q-gram matches.

In total, I was able to match the following:

Classification	Observations	Unique Employees
Contractor	1328	1028
Civil Servant	1658	1218
Individual not found	463	375
Neither data source	10	9
No contractor data	40	34
No civil servant data	121	97
Proportion Identified	0.825	0.813
Proportion Identified | ≥ 1 Call answered in municipality	0.831	0.821
Proportion Identified | Administrative Data	0.843	0.843
Proportion Identified | Call data	0.808	0.789

Table A.12: Matched sample of bureaucrats from the calls and program officer data.

Government interviews suggested that the data quality is worse in small and poor municipalities. Based on their comments, I analyze the correlates of missingness/unidentifiability in the data briefly showing that the highest rates of unidentifiability occur in very poor municipalities. In Figure A.10 I examine an index of municipal poverty (2005 census) and estimated 2018 population as predictors of missingness. Unidentifiability is increasing (monotonically) in the poverty rate in each municipality.

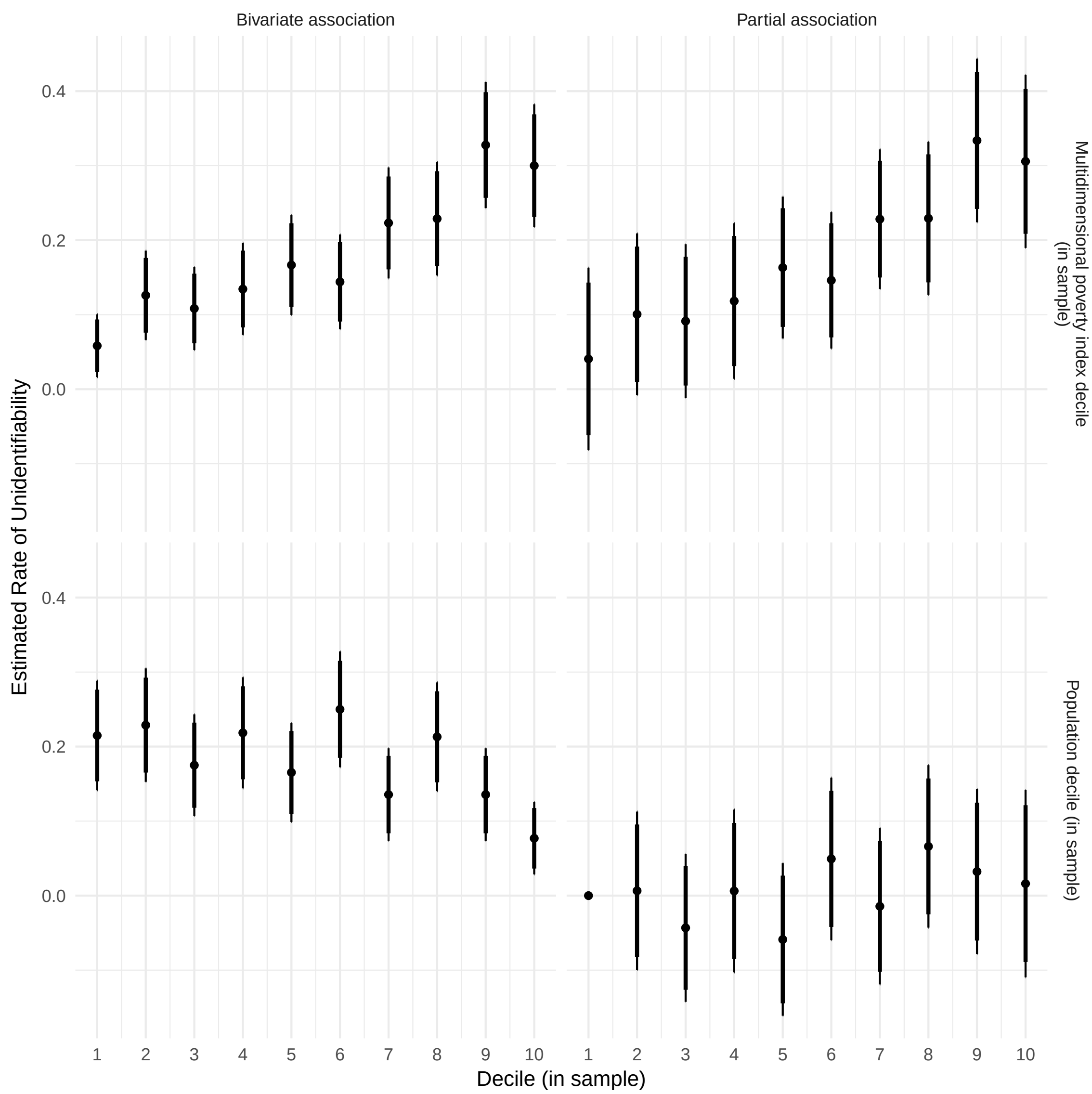


Figure A.10: The association between the DANE multidimensional municipal poverty index (2005), estimated population (2018), and failure to identify employees in the database. The left column is estimated from two separate bivariate regressions. The right column depicts partial associations from a model with both sets of indicators. Confidence intervals are estimated on the basis of heteroskedasticity-robust standard errors.

A.12.3 Individual Measures

The main individual measure is an indicator of civil service versus patronage. Within each subset, I have some covariates (with varying degrees of missingness). For the purposes of contracts, I measure the tenure of the individual's employment as the date since the first contract was signed. I also note whether this occurred before or after the last transition of municipal government (January 1, 2016).

A.12.4 Municipal-Level Measures

At the municipal level, I measure two variables from the contract data:

- Median length of contracts, in days. Principals can theoretically exercise more control by holding bureaucrats to shorter contracts.

- Degree of turnover following the last municipal transition of power. I examine the degree of turnover after the last change of municipal government in 2016. This quantity is the share of contractors hired during 2016 and 2017 that had previously served as a contractor in public administration in the *alcaldía*. Stylized examples from Bucaramanga and Cúcuta two large cities in the same region illustrate the difference in proportion graphically in Figure A.11. A much larger share of Bucaramanga's employees carried on after the election.

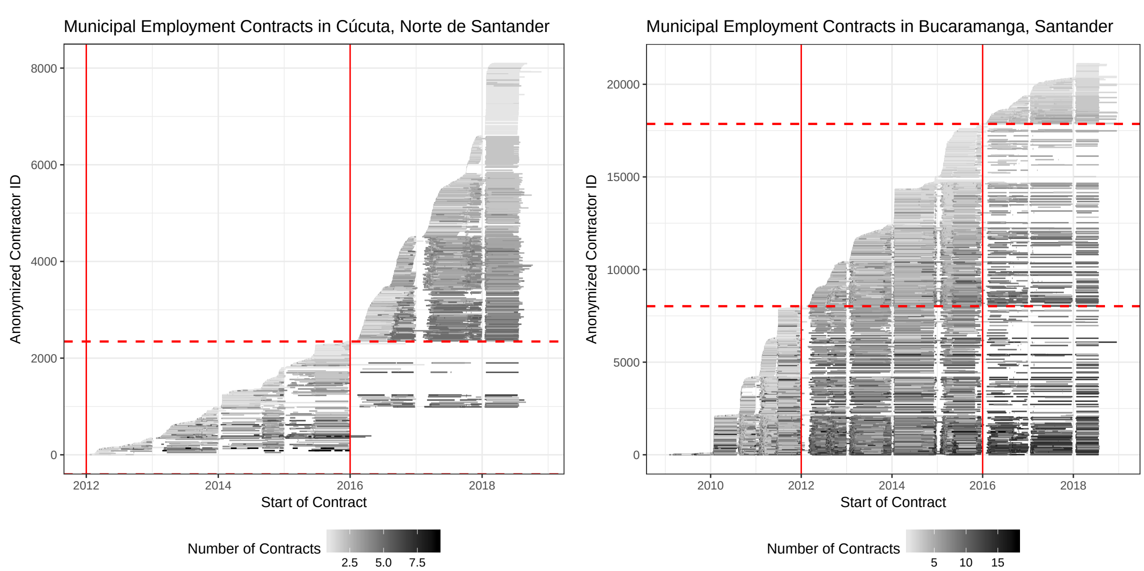


Figure A.11: All contracts in SECOP-I for Cúcuta and Bucaramanga, Colombia.

A.13 Demographic Data

The demographic data, measured at the municipal level, that is used in analyses come from:

- Population projections as of 2018. I use DANE's 2018 population projections for municipalities and the Alcaldía Mayor's projections of Bogotá's population, by locality, in 2018.
- Census of 2005: Municipal education levels and municipal multidimensional poverty index.
- Census microdata: Rates of migration are obtained from all census microdata available through IPUMS.
- SISBÉN registrations: Data downloaded from SISBÉN's open data as of November, 2017. Available at `https://www.sisben.gov.co/Territorios/Paginas/Reportes%20Base%20Certificada/ano2018.aspx`.
- Population of internally displaced persons (IDPs), by municipality. I obtain data on the number of IDPs per municipality from Unidad de Victímas as of July 1, 2018. Available at `http://cifras.unidadvictimas.gov.co/Home/Desplazamiento`.

A.14 Election Data

Election data came from the dataset at Universidad de los Andes. This data is compiled from the Registraduría Nacional del Estado Civil by the Centro de Estudios sobre Desarollo Económico. All candidates in municipal elections are available since 1997. I examine on the municipal elections of 1997, 2000, 2003, 2007, 2011, and 2015.

Extreme weakness of parties limited serial correlation render standard measures of competitiveness poor measures. The correlation between mayoral margins of victory at time $t-1$ and t are depicted in Figure A.12.

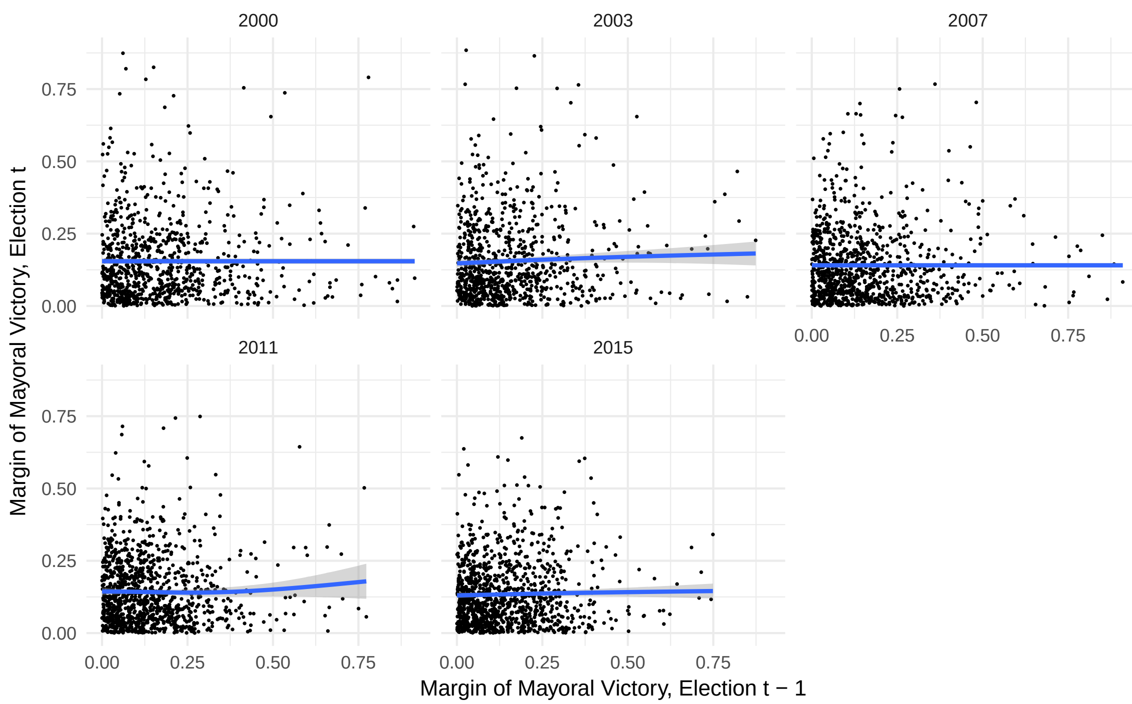


Figure A.12: There is effectively no correlation between mayoral margin of victory at time $t-1$ and time t. Each point represents a municipality. The lines are fit with GAM.

Instead I use three measures of municipal political competition from both local mayoral and council elections.

- Mayoral election effective number of candidates (ENC). This measure is increasing in competitiveness. It is calculated via the following formula, where i index candidates and p_i is each candidate's vote share.

$$ENC = \frac{1}{\sum_{i=1}^{n} p_i^2}$$

- Number of unique council members since 1997. Council members are not term limited. This is calculated by the following formula.

$$\text{Ratio of Unique Concejales} = \frac{\text{\# Unique winners, 1997-2015}}{\text{\# Council seats, 1997-2015}}$$

- The Colombian naming tradition is to give a child two last names. The first is the father's first last name; the second is the mother's first last name. Most candidates use both last names. To measure concentration of power by family, I examine the ratio of unique last names to total last names ($\approx 2\times$ number of candidates).

$$\text{Ratio of Unique Last Names} = \frac{\text{\# Unique last names of concejales, 1997-2015}}{\text{\# Total last names of consejales, 1997-2015}}$$

A.15 Sample Selection

As indicated in Table 3, I estimate the bias outcomes on the sample of answered calls. While the blocking scheme ensures balance across municipalities and within the sample of calls, it is worth assessing whether or not the resultant sample is imbalanced across the experimental factors. This is assessed through an F-test of the joint significance of the four unrevealed factors: class, accent, migrant status, and question difficulty.

I show in Table A.13 that there is no evidence that entering this sample (Columns 3-4) is endogenous to the unrevealed factors. The test of this claim is the F-test on the placebos, for each estimator. Further, Columns 1-2 and 5 show that the unrevealed factors do not predict how soon or late a call was answered. In sum, there is no evidence that answering is endogenous to any of the analyzed treatments.

	Did not Answer First Call		Did not Answer Any Calls		Number of Calls
	(1)	(2)	(3)	(4)	(5)
Afternoon Call	0.031*	0.022	0.039**	0.030**	0.291*
	(0.023)	(0.018)	(0.023)	(0.015)	(0.188)
DV mean, Morning	0.639	0.639	0.330	0.330	4.48
Estimator	OLS	OLS	OLS	OLS	Tobit
IPW	✓		✓		✓
Entity, Enumerator FE		✓		✓	
Program Indicator	✓	✓	✓	✓	✓
Placebos	✓	✓	✓	✓	✓
Placebo F-test*, p-value	0.898	0.381	0.705	0.163	0.619
Hypothesis test	Upper	Upper	Upper	Upper	Upper
DV range	{0, 1}	{0, 1}	{0, 1}	{0, 1}	[1,7]
Censoring					Right
Observations	1,836	1,836	1,836	1,836	1,836

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.13: The AMCEs of an afternoon call on measures of absenteeism of the dispatcher. Six calls were attempted over the course of at least three days; Columns 3-4 indicate that none of the six calls were answered. Heteroskedasity-robust standard errors for the OLS models in parentheses. *For the tobit model, the p-value on the test of joint significance of the placebos is calculated from a Pearson χ^2 likelihood ratio test. The mean outcome in the morning in the fifth column refers to the latent outcome and is estimated by tobit.

A.16 Robustness of Bias Estimates

A.16.1 Complier AMCE Estimates for the Class Treatment

The results on the class treatment are central to findings of bias. There are two main concerns about this treatment, with responses below.

1. Can people judge a caller's class by voice alone?

   - The results from the double coding of compliance in Table A.8 suggest as much: 77.5% of calls were correctly identified; 13.5% were indeterminate; and only 9% were opposite of the assignment indicator.

2. Class is a compound treatment in any case in Colombia. The scripts that distinguish the

classes include different salutations, different vocabularies, and different presentation of the question.

- The pervasiveness of social class in the Colombian context is important; isolating class from its correlates or constituent parts is not particularly feasible, nor is it particularly useful in this context.
- The results show that class-based bias happens in certain processes and not others. There is no evidence that bureaucrats did not understand questions from poor petitioners, as rates of correct responses are not substantially different between lower- and lower-middle class Colombians. These facts suggest that there was no systematic response to the characteristics used to connote class.

To test further concerns about the excludability of treatment assignment in Table A.14, I estimate complier AMCEs on the class treatment, instrumenting an observed lower-middle class exchange with assignment to the lower-middle class treatment condition. If results are driven by perceptions of class and coder ratings of class are correlated with how bureaucrats perceived class, then (non-zero) point estimates of complier AMCEs should be higher than intent-to-treat AMCEs. I report the class estimates from Table 4 (Panels A and B) in the main text along their complier analogues.

Note that this test serves as an informal test of excludability. It is also possible that if some characteristic driving the results that is highly positively correlated with observer judgments of class is driving bias in behavior, the complier estimates would be higher than the intent-to-treat estimates. All of the point estimates on the outcomes where there is evidence for bias in the ITT AMCE estimates in Table 4 in the main text increase in magnitude.

	Access/Process				Response to Petition					
	Dispatcher Gave Name	Petition Made	Second Official	Program Officer	Complete	Incomplete	Any Info.	Red Tape	*Alcaldía* Only	extra
PANEL A: INTENT-TO-TREAT AMCE ESTIMATES, IPW ADJUSTMENT										
Lower-Middle Class	0.011	0.014	0.017	0.044	0.021	0.048*	0.069**	−0.019	−0.036**	0.003
	(0.020)	(0.025)	(0.029)	(0.028)	(0.018)	(0.029)	(0.029)	(0.021)	(0.017)	(0.024)
PANEL B: COMPLIER AMCE ESTIMATES, IPW ADJUSTMENT										
Lower-Middle Class	0.028	0.019	0.033	0.069	0.035	0.065	0.100**	−0.028	−0.053**	0.003
	(0.028)	(0.037)	(0.044)	(0.042)	(0.027)	(0.043)	(0.043)	(0.031)	(0.025)	(0.036)
PANEL C: INTENT-TO-TREAT AMCE ESTIMATES, ENTITY FIXED EFFECTS										NA
Lower-Middle Class	−0.015	0.016	−0.004	0.049*	0.024	0.057**	0.081***	−0.027	−0.039**	0.002
	(0.019)	(0.025)	(0.028)	(0.025)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
PANEL D: COMPLIER AMCE ESTIMATES, ENTITY FIXED EFFECTS										
Lower-Middle Class	0.002	0.028	0.009	0.083**	0.038	0.087**	0.125***	0.006	−0.055**	0.006
	(0.025)	(0.036)	(0.041)	(0.037)	(0.028)	(0.040)	(0.040)	(0.037)	(0.025)	(0.037)
Mean DV, Paisa Accent	0.842	0.723	0.53	0.296	0.119	0.414	0.533	0.124	0.066	0.266
All Factors	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DV range	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}
Observations	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.14: Estimates of the intent-to-treat and complier AMCEs of the class treatment. Heteroskedasticity robust standard errors in parentheses. "Observed" indicates that the accent was observed by double coders that listened to all call recordings.

A.16.2 Regional Accent and Red Tape Robustness Test

Table 4 documents the disproportionate rate at which red tape was demanded from petitioners with a Paisa regional accent. Because the regional accent did not vary within an enumerators' calls, I conduct tests that drop enumerators one by one as well as in trios (one enumerator per accent) to examine the robustness of this finding. Figures A.13 and A.14 show the point estimates when dropping one enumerator and one trio of enumerators, respectively. They suggest that the inference is generally robust and that the effect does not seem to be driven by any single enumerator or pair/trio of enumerators.

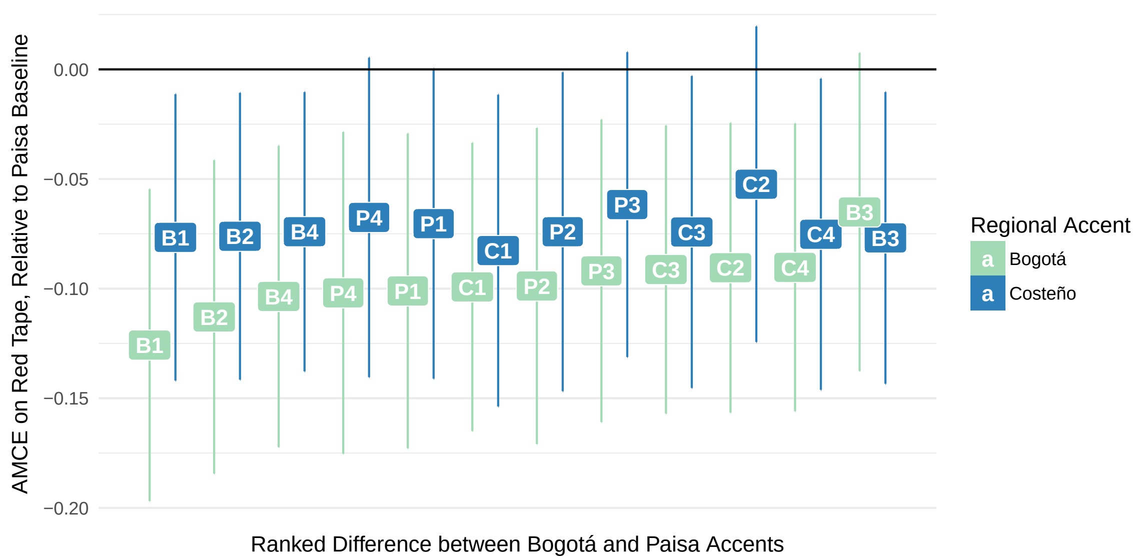


Figure A.13: AMCE of Bogotá and Costeño accents relative to a Paisa accent when dropping one enumerator. The enumerator codes are labeled on the points and indexed by accent ("B", "C", or "P") and number. The bars indicate 95% confidence intervals calculated on heteroskedasticity-robust standard errors.

A.16.3 Regional Accents in Home Region

In the main results, I analyze regional accents without regard to the match between the accent of the petitioner and the accent native to the municipality where the audit was conducted. This analysis redefines the treatment indicator for accent as being an "in-region" (home) accent. Because the

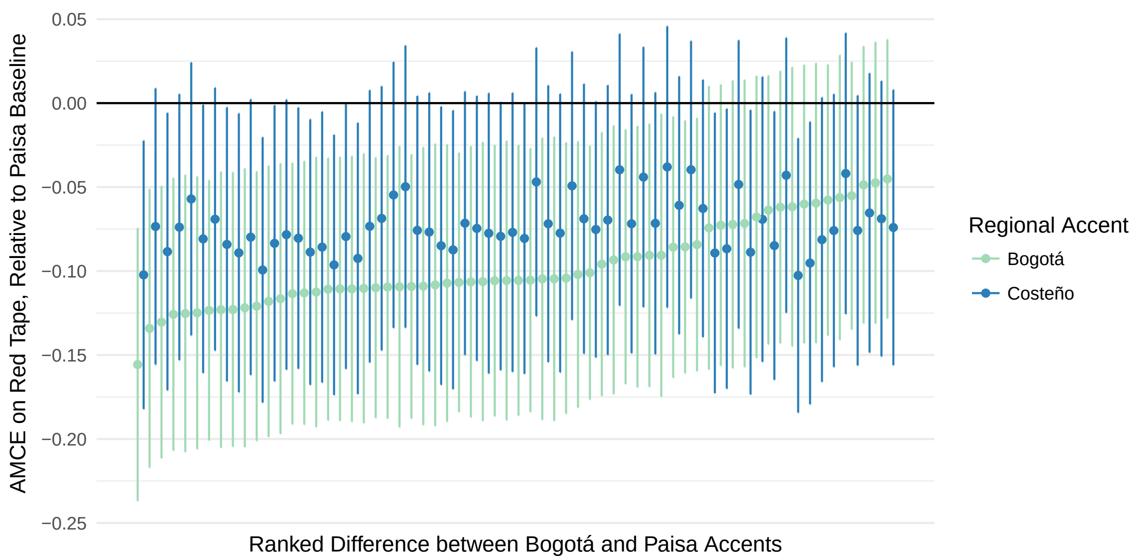


Figure A.14: AMCE of Bogotá and Costeño accents relative to a Paisa accent when dropping each permutation of three enumerators (one per accent). Note that by dropping three enumerators simultaneously, the effective sample is approximately 75% of the main sample reported in Table 4, inflating standard error estimates accordingly. The bars indicate 95% confidence intervals calculated on heteroskedasticity-robust standard errors.

experimental accent treatment only includes 3 of $\approx$ 12 regional accents, I subset to the regions of Colombia in which there is common support for the treatment. Table A.15 clarifies the definition of the region for purposes of analysis as well as the total number of petitions in the subsample. In sum, this sample represents about 55% of all answered calls.

With both adjustment strategies (IPW and entity fixed effects), the region indicator is interacted with all factors and an indicator for the program. Results are reported in Table A.16. The main takeaways are as follows:

- We cannot reject the null hypothesis that the conditional AMCE = 0 for the pooled subsample or any subgroup therein.

- For some outcomes, there are statistically significant differences between regions. If anything, these differences seem to be driven by the fact that, Costeño confederates seem to receive

				n of Petitions	
Accent	Region	Departments	Total	"In-region"	"Out-region"
Bogotá	Centro Oriente (subset)	(Bogotá) Cundinamarca	198	65	133
Costeño	Caribe	Atlántico Bolívar Cesar Córdoba La Guajira Magdalena San Andrés Sucre	189	58	131
Paisa	Eje Cafetero - Antioquia	Antioquia Caldas Quindío Risaralda	270	85	185

Table A.15: Definition of regions for the analysis of "in-region" accents. The sample from which these municipalities are drawn is the 1194 answered calls. Note that by construction, $\frac{1}{3}$ of calls should be "in-region"; this proportion is maintained in this subsample.

slightly "worse" service "in region" in the Atlantic Coast (Caribe). This was not anticipated.

A.16.4 Migrant Status in the "Call Answered" and "Petitioned" Samples

Migrant status was not revealed until petitions were made (see Table 3), however the main analysis analyzes outcomes based on migrant status within the sample of all answered calls (Table 6). This serves to inflate the effective rate of noncompliance for the migrant factor. This attenuates the resultant intent-to-treat estimates. The F-tests in Tables 4 and 6 (columns 1-4) suggest that there is no reason to believe that migrant status was revealed prior to the petition. As such, the estimates in the smaller sample should be larger in magnitude than those in the main text. Table A.17 supports this interpretation.

A.16.5 Survey Outcomes

In this section, I examine the responses of enumerators to survey questions about their experience of each call. The survey consisted of five questions, translated to English as follows:

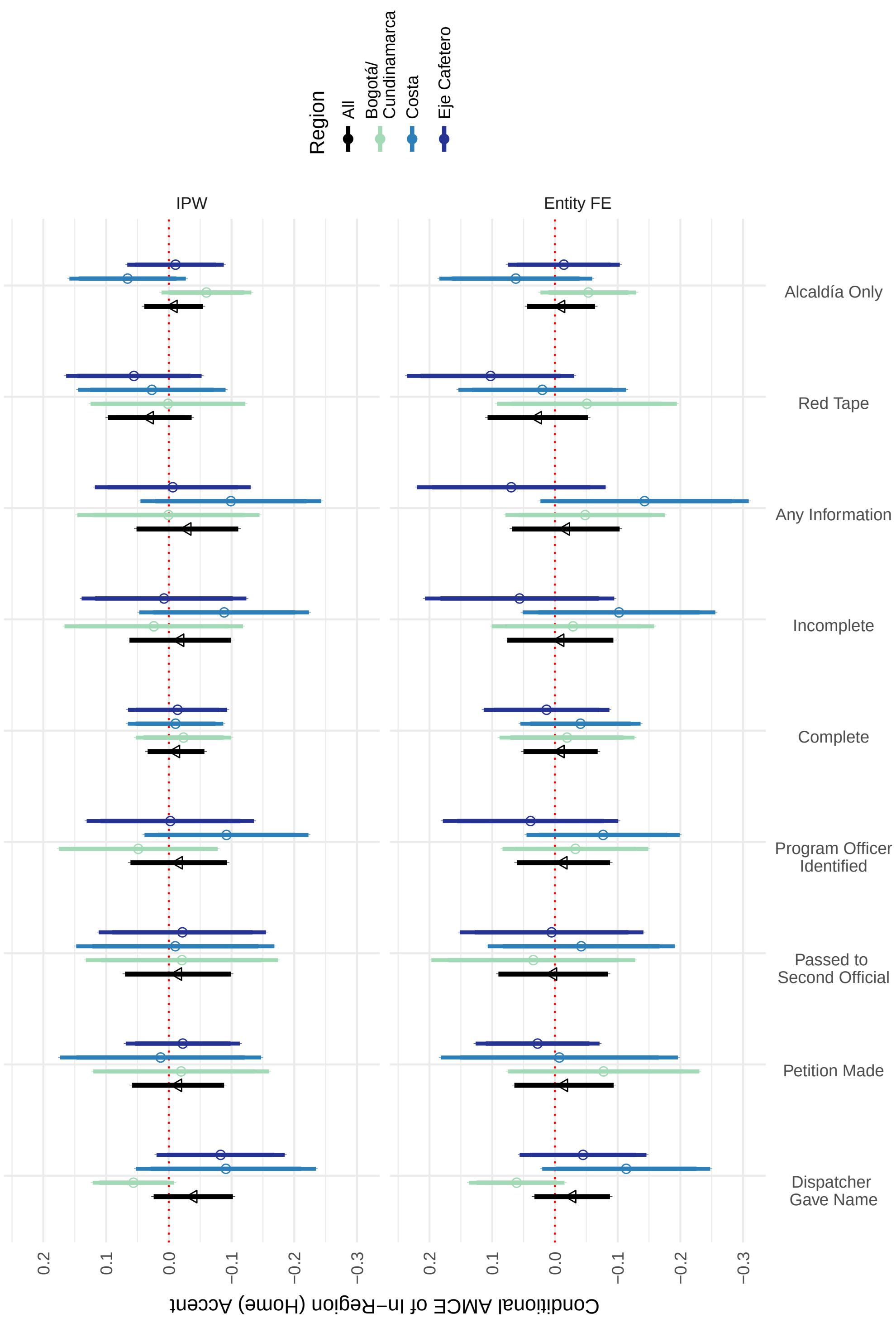


Table A.16: Estimates of the conditional AMCE of an "in region" accent petition. Bars represent 95% confidence intervals constructed on heteroskedasticity robust standard errors. The panels correspond to the adjustment strategy used in estimation.

	Response to Petition									
	Complete		Incomplete		Any Info.		Red Tape		*Alcaldía* Only	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A: IPW ESTIMATES										
Migrant	−0.002	−0.003	−0.027	−0.054	−0.029	−0.057*	−0.036	−0.049	0.054***	0.070***
	(0.017)	(0.021)	(0.029)	(0.033)	(0.029)	(0.030)	(0.024)	(0.030)	(0.017)	(0.021)
PANEL B: ESTIMATES WITH ENTITY FIXED EFFECTS										
Migrant	0.003	−0.003	−0.044	−0.065*	−0.041	−0.067**	−0.041	−0.033	0.061***	0.081***
	(0.018)	(0.025)	(0.028)	(0.035)	(0.028)	(0.032)	(0.026)	(0.034)	(0.017)	(0.023)
PANEL C: ESTIMATES WITH ENTITY + ENUMERATOR FIXED EFFECTS										
Migrant	0.002	0.0003	−0.043	−0.063*	−0.041	−0.063**	−0.040	−0.029	0.061***	0.077***
	(0.018)	(0.026)	(0.028)	(0.035)	(0.028)	(0.032)	(0.026)	(0.035)	(0.017)	(0.023)
Mean DV, Resident	0.102	0.135	0.447	0.592	0.731	0.713	0.252	0.334	0.061	0.081
Program	✓	✓	✓	✓	✓	✓	✓	✓	✓	
All Factors	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DV range	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}
Sample	Answered	Petitioned	Answered	Petitioned	Answered	Petitioned	Answered	Petitioned	Answered	Petitioned
Observations	1,194	911	1,194	911	1,194	911	1,194	911	1,194	911

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.17: Estimates of the AMCEs of the migrant treatment on both the call answered subsample from the main text as well as the petitioned sample. Heteroskedasticity robust standard errors in parentheses.

1. *Satisfaction:* On a scale from 1 to 5, where 1 is "not at all satisfied" and 5 is "very satisfied," how satisfied are you with the service given by the public servant during the call?

2. *Confidence (in answers):* On a scale from 1-4, where 1 is "not confident" and 4 is "very confident," how much confidence did you have in the responses given by the public servant?

3. *Actionable information:* On a scale from 1 to 5, where 1 is "very easy" and 5 is "very difficult," how hard would it be to carry out the process (service) you asked about on the basis of the information you received?

4. *Knowledge:* On a scale from 1 to 5, where 1 is "very low" and 5 is "very high," what level of knowledge did the public servant have when responding to the request?

5. *Respect:* On a scale from 1 to 5, where 1 is "very little respect" and 5 is "lots of respect," how respectful was the public servant while responding to the request?

Due to an issue in the programming of the survey, responses are missing for 59/1194 answered calls and 30/911 of the calls in which a petition was made. In these cases, the survey did not appear at the end of data input. The differential proportions of missing survey responses indicate that this error occurred as part of the sequence in the survey. This may be endogenous to some of the experimental manipulations, if through the trajectory of the call. However, missingness is balanced across the factors.

My main measure of service provision is a z-score index comprising the five measures enumerated above. Component # 3 is reversed such that higher scores on the scale map onto higher values of the index (better service). I also report the standardized measure of respect given arguments made in the paper. Estimates of AMCES of the experimental manipulations on these outcomes are reported in Table A.18.

Note that the enumerator effects reported seem to correspond to idiosyncracies in how individual enumerators assess service. The estimates are *not* robust to dropping one enumerator at a time. I focus on the within-enumerator estimates in columns (3) and (6) of both panels as the main measure of the relationship between petition and petitioner attributes and service provision.

A.17 Supporting Tables for Tests of the Mechanism

This section provides tables to support the graphical analysis in Section 6 of the main text. The first tests examine the cost of effort and employee type (contractor/civil servant). For covariates with within-*alcaldía* variation, I do not run the specification with *alcaldía* fixed effects, since this reduces the effective sample dramatically. Instead, I run a second specification with indicators for stratum (to account for differential probabilities of assignment) and enumerator.

A.17.1 Distinguishing Bureaucrat Taste-Driven from Oversight-Driven Bias

Table A.19 shows that bias in information provision by class was attenuated to zero on technical questions. While the interaction is not generally significant, differences are quite stark. Note, however, that the theory implies a one-tailed test while the table reflects (conservative) two-tailed p-values.

	Dependent variable:					
	Service Index (Standardized)			Respect (Standardized)		
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: CLASS AND ACCENT ON ANSWERED CALLS SUBSAMPLE						
Lower Middle Class	0.109**	0.110**	0.096**	0.020	0.010	0.002
	(0.047)	(0.047)	(0.045)	(0.059)	(0.061)	(0.058)
Bogotá Accent	0.287***	0.226***		0.238***	0.191**	
	(0.058)	(0.063)		(0.075)	(0.081)	
Costeño Accent	0.062	0.064		−0.297***	−0.305***	
	(0.059)	(0.063)		(0.079)	(0.084)	
DV mean, Lower Class	-0.052	-0.052	-0.052	-0.006	-0.006	-0.006
DV mean, Paisa	-0.106	-0.106	-0.106	0.033	0.033	0.033
DV Range	[-2.957, 1.904]	[-2.957, 1.904]	[-2.957, 1.904]	[-3.767, 1.263]	[-3.767, 1.263]	[-3.767, 1.263]
Observations	1,135	1,135	1,135	1,135	1,135	1,135
PANEL B: MIGRANT STATUS AND QUESTION DIFFICULTY ON PETITION SUBSAMPLE						
Migrant	0.060	0.085	0.085	0.109*	0.143**	0.125*
	(0.049)	(0.057)	(0.055)	(0.061)	(0.072)	(0.071)
Technical Petition	−0.128**	−0.110*	−0.117**	−0.124**	−0.107	−0.114
	(0.050)	(0.058)	(0.056)	(0.063)	(0.075)	(0.072)
DV mean, Easy	-0.052	-0.052	-0.052	-0.006	-0.006	-0.006
DV mean, Resident	0.062	0.062	0.062	-0.003	-0.003	-0.003
DV Range	[-2.957, 1.904]	[-2.957, 1.904]	[-2.957, 1.904]	[-3.767, 1.263]	[-3.767, 1.263]	[-3.767, 1.263]
Observations	881	881	881	881	881	881
Estimator	IPW	Entity FE	Entity, Enum. FE	IPW	Entity FE	Entity, Enum. FE
Program	✓	✓	✓	✓	✓	✓
All Factors	✓	✓	✓	✓	✓	✓
Note:						*p<0.1; **p<0.05; ***p<0.01

Table A.18: The AMCEs of treatment conditions on enumerators' perceived treatment and service by bureaucrats. The outcome for columns 1-3 (both panels) is a Z-score index of five attributes of service. The outcome for columns 4-6 is the standardized "respect" component of that index. Heteroskedasticity-robust standard errors in parentheses.

	Complete	Incomplete	Any Info.	Alcaldía Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
PANEL A: CONDITIONAL AMCE BY DIFFICULTY OF PETITION; IPW ESTIMATES					
Lower-Middle Class	0.039	0.080**	0.118***	−0.040***	0.036
	(0.026)	(0.038)	(0.038)	(0.015)	(0.034)
Hard: Lower-Middle Class	−0.034	−0.067	−0.101*	0.009	−0.064
	(0.035)	(0.057)	(0.058)	(0.034)	(0.048)
Conditional Effect, Technical Petition	0.004	0.013	0.017	-0.031	-0.029
	(0.023)	(0.042)	(0.044)	(0.030)	(0.034)
PANEL B: CONDITIONAL AMCE BY DIFFICULTY OF PETITION; STRATUM + ENUMERATOR FE					
Lower-Middle Class	0.030	0.068*	0.098***	−0.037**	0.030
	(0.026)	(0.037)	(0.038)	(0.015)	(0.034)
Hard: Lower-Middle Class	−0.021	−0.049	−0.070	0.004	−0.054
	(0.034)	(0.056)	(0.057)	(0.034)	(0.048)
Conditional Effect, Technical Petition	0.010	0.019	0.029	-0.033	-0.024
	(0.023)	(0.042)	(0.043)	(0.030)	(0.033)
Observations	1,194	1,194	1,194	1,194	1,194
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A.19: Estimates of the conditional AMCE of a lower-middle class petitioner by the difficulty of the question. The base category is the "easy"/inscriptions questions. The experimental treatment technical petition is interacted with all factors and a program indicator. "Conditional effect" refers to the conditional effect of lower-middle class. Heteroskedasticity- robust standard errors in parentheses.

Table A.20 shows that bias in information provision by class was attenuated to zero for MFA petitions. While the interaction is not generally significant, differences are large and robust to different estimators. Note that the theory implies a one-tailed test while the table reflects (conservative) two-tailed p-values.

A.17.2 Isolating Complaint-Driven Bias

Table A.21 examines the conditional effect of class by level of municipal poverty. Panels A and B show the estimates using both of the main estimators in the manuscript. Panels C and D demonstrate that these findings are robust to the use of flexible, interactive controls for municipal population.

	Complete	Incomplete	Any Info.	Alcaldía Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
PANEL A: CONDITIONAL AMCE BY DIFFICULTY OF PETITION; IPW ESTIMATES					
Lower-Middle Class	0.003	0.030	0.033	−0.021	0.019
	(0.021)	(0.040)	(0.041)	(0.020)	(0.031)
SISBÉN: Lower-Middle Class	0.036	0.031	0.067	−0.029	−0.034
	(0.035)	(0.057)	(0.058)	(0.033)	(0.049)
Conditional Effect, SISBÉN	0.039	0.061	0.100**	-0.050*	-0.015
	(0.028)	(0.041)	(0.041)	(0.026)	(0.037)
PANEL B: CONDITIONAL AMCE BY DIFFICULTY OF PETITION; STRATUM + ENUMERATOR FE					
Lower-Middle Class	0.007	0.022	0.029	−0.019	0.019
	(0.020)	(0.039)	(0.040)	(0.019)	(0.031)
SISBÉN: Lower-Middle Class	0.028	0.047	0.075	−0.032	−0.029
	(0.035)	(0.056)	(0.057)	(0.031)	(0.048)
Conditional Effect, SISBÉN	0.035	0.069*	0.104***	-0.051**	-0.010
	(0.029)	(0.040)	(0.040)	(0.025)	(0.037)
Observations	1,194	1,194	1,194	1,194	1,194
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A.20: Estimates of the conditional AMCE of a lower-middle class petitioner by the audited program (MFA or SISBEÉN). The base category is MFA. The program indicator is interacted with all factors in the experimental design. “Conditional effect” refers to the conditional effect of lower-middle class. Heteroskedasticity-robust standard errors in parentheses.

One alternative explanation for the positive association between class-bias and poverty rate is clientelism. I conduct three tests to show that the evidence is inconsistent with this explanation. First, I show that levels of service given to the lower middle class do not vary in municipal poverty rates. If a politician were simply co-opting a social program to devote services clentelistically as in Weitz-Shapiro (2012), we would expect lower levels of service by bureaucrats across the board (e.g. even for the lower middle class). If this happens disproportionately in poor places, then there should be a negative association between municipal poverty rates and service outcomes for the lower middle class. Table A.22 indicates that this is not the case. There is little evidence of a correlation,

	Complete (1)	Incomplete (2)	Any Info. (3)	Alcaldía Only (4)	Red Tape (5)
PANEL A: CONDITIONAL AMCE BY TERCILE OF POVERTY RATE; IPW ESTIMATES					
Lower-Middle Class	0.013	−0.024	−0.011	−0.027	−0.007
	(0.029)	(0.046)	(0.051)	(0.028)	(0.045)
Medium Poverty: Lower-Middle Class	0.019	0.085	0.105	−0.025	0.062
	(0.042)	(0.067)	(0.070)	(0.043)	(0.062)
High Poverty: Lower-Middle Class	0.001	0.139^{**}	0.139^{**}	−0.012	−0.032
	(0.045)	(0.066)	(0.070)	(0.040)	(0.060)
Conditional Effect in M	0.032	0.061^{*}	0.094^{**}	-0.051^{*}	0.055^{*}
	(0.028)	(0.047)	(0.046)	(0.034)	(0.039)
Conditional Effect in H	0.014	0.115^{***}	0.128^{***}	-0.038^{*}	-0.039
	(0.035)	(0.045)	(0.044)	(0.026)	(0.040)
PANEL B: CONDITIONAL AMCE BY TERCILE OF POVERTY RATE; ENTITY + ENUMERATOR FE					
Lower-Middle Class	0.019	−0.024	−0.005	−0.020	−0.013
	(0.036)	(0.050)	(0.053)	(0.030)	(0.054)
Medium Poverty: Lower-Middle Class	0.025	0.091	0.117	−0.039	0.080
	(0.051)	(0.077)	(0.077)	(0.050)	(0.074)
High Poverty: Lower-Middle Class	−0.001	0.151^{**}	0.150^{**}	−0.030	−0.033
	(0.057)	(0.077)	(0.076)	(0.048)	(0.075)
Conditional Effect in M	0.044	0.067	0.111^{**}	-0.058^{*}	0.067^{*}
	(0.037)	(0.061)	(0.060)	(0.042)	(0.054)
Conditional Effect in H	0.018	0.127^{**}	0.145^{***}	-0.050^{*}	-0.047
	(0.048)	(0.060)	(0.059)	(0.039)	(0.056)
PANEL C: CAMCE BY TERCILE OF POVERTY RATE WITH INTERACTIVE POPULATION CONTROL; IPW ESTIMATES					
Lower-Middle Class	0.010	0.019	0.029	−0.033	0.004
	(0.036)	(0.056)	(0.056)	(0.036)	(0.058)
Medium Poverty: Lower-Middle Class	0.028	0.037	0.065	−0.012	0.054
	(0.052)	(0.080)	(0.081)	(0.056)	(0.077)
High Poverty: Lower-Middle Class	−0.009	0.111	0.102	−0.005	−0.034
	(0.053)	(0.082)	(0.082)	(0.057)	(0.083)
Conditional Effect in M	0.039^{*}	0.056	0.095^{**}	-0.045^{*}	0.057^{*}
	(0.031)	(0.050)	(0.047)	(0.036)	(0.040)
Conditional Effect in H	0.001	0.130^{***}	0.131^{***}	-0.038	-0.030
	(0.036)	(0.048)	(0.049)	(0.032)	(0.048)
Interactive Population Decile Bins	✓	✓	✓	✓	✓
PANEL D: CAMCE BY TERCILE OF POVERTY RATE WITH INTERACTIVE POPULATION CONTROL; ENTITY + ENUMERATOR FE					
Lower-Middle Class	0.023	−0.003	−0.011	−0.030	−0.020
	(0.045)	(0.069)	(0.062)	(0.041)	(0.072)
Medium Poverty: Lower-Middle Class	0.028	0.054	0.113	−0.021	0.087
	(0.065)	(0.096)	(0.083)	(0.062)	(0.094)
High Poverty: Lower-Middle Class	−0.028	0.146	0.159^{*}	−0.017	−0.021
	(0.071)	(0.105)	(0.093)	(0.067)	(0.107)
Conditional Effect in M	0.051^{*}	0.051	0.101^{**}	-0.051	0.067^{*}
	(0.040)	(0.065)	(0.060)	(0.045)	(0.053)
Conditional Effect in H	-0.005	0.144^{**}	0.147^{**}	-0.048	-0.041
	(0.049)	(0.065)	(0.063)	(0.044)	(0.065)
Interactive Population Decile Bins	✓	✓	✓	✓	✓
Observations	1,194	1,194	1,194	1,194	1,194

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.21: Estimates of the conditional AMCE by tercile of poverty. The base category is the first tercile (lowest rate of poverty). All covariates and moderators (poverty tercile indicator and demeaned poverty decile bins) are interacted across all factors in the design. Standard errors are clustered at the municipality level since this is the level of measurement of the poverty moderator.

and if anything, there is weak evidence of a a *positive* correlation between municipal poverty and information provision. This correlation is not robust to alternate functional forms (Panel B) or to dropping Bogotá (Panels C-D), which accounts for a disproportionate share of observations. This provides no evidence in favor of a clientelism explanation for observed findings.

Second, I draw on documentation of clientelism in Colombia to identify variation in the presumed intensity of clientelism across municipalities. I show that the positive association between class-bias and poverty rate is robust to controlling interactively for these measures. The tests are as follows. First, I show that there is no association between documented threats to electoral integrity. These threats include clientelism, fraud, intimidation, and electoral violence. To the extent that an emerging literature on clientelism suggests that clientelism includes both "carrots" and "sticks," this provides a measure of both instruments (Mares and Young, 2018). Using data from the Misión de Observacion Electoral in Colombia, I control for a binary indicator for a general predicted threat to the 2018 national elections (Misión de Observación Electoral, 2018). Where a threat is identified, I code this variable as a "1." This creates two categories, which I demean and interact across all factors and the program indicator. Panels A and B of Table A.23 suggest that results are not sensitive to estimating effects within levels of electoral threat.

Writing on clientelism in Colombia suggest that clientelism is practiced in distinct patterns in different regions (Ocampo, 2014). To account for these patterns, I include interactive department ($n = 30$) fixed effects in Panels C and D of Table A.23. Note that some departments have few municipalities and few calls, so Panel D, in particular represents a subset of the sample in departments where there is variation in both population category within department. Nevertheless, results are consistent with the broader patterns documented in the main text and in Table A.21. These analyses

	Complete	Incomplete	Any Info.	Alcaldía Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
Panel A: Linear Association between Poverty and Outcomes, Lower-Middle Class Petitioners Only					
Poverty rate	−0.011	0.268***	0.255**	−0.017	−0.098
	(0.060)	(0.104)	(0.105)	(0.043)	(0.080)
Observations	600	600	600	600	600
Panel B: Quadratic Association between Poverty and Outcomes, Lower-Middle Class Petitioners Only					
Poverty rate	0.195	0.775	0.970	0.445	−0.091
	(0.360)	(0.616)	(0.593)	(0.283)	(0.499)
Poverty rate^2	−0.182	−0.449	−0.631	−0.409*	−0.006
	(0.311)	(0.535)	(0.508)	(0.238)	(0.428)
Observations	600	600	600	600	600
Panel C: Linear Association between Poverty and Outcomes, Lower-Middle Class Outside of Bogotá					
Poverty rate	−0.008	0.053	0.044	−0.028	−0.136
	(0.066)	(0.107)	(0.593)	(0.047)	(0.091)
Observations	559	559	559	559	559
Panel D: Quadratic Association between Poverty and Outcomes, Lower-Middle Class Outside of Bogotá					
Poverty rate	0.264	−0.747	−0.483	0.477	−0.379
	(0.419)	(0.645)	(0.591)	(0.311)	(0.586)
Poverty rate^2	−0.233	0.685	0.452	−0.432*	0.208
	(0.350)	(0.554)	(0.505)	(0.255)	(0.484)
Observations	559	559	559	559	559
Factors (not Class)	✓	✓	✓	✓	✓
Program Indicator	✓	✓	✓	✓	✓
Estimator	IPW	IPW	IPW	IPW	IPW
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A.22: Estimates of the association between municipal poverty rates and service devoted to the lower middle class. The sample includes only calls made by lower-middle class petitioners, and in Panels C and D, only calls made outside of Bogotá. All covariates and and a program indicator are interacted across all factors in the design. Standard errors are clustered at the municipality level since this is the level of measurement of poverty rates.

provide no evidence that clientelism is driving the observed association between poverty and bias.

Figure A.15 shows no evident variation in class-based bias as a function of local political competition. The three columns use three alternate measures of local political competition. The left column uses the ratio of unique councilors to total councilors (1997-2015); the middle column uses the effective number of mayoral candidates in the last 3 elections; and the last column uses an inverse covariance weighted index of the first two plus unique last names (*apellidos*) over council elections from 1997-2015.

A.18 Robustness of Link to Administrative SISBÉN Enrollment

For the analysis of SISBÉN enrollment and bias in Section 7, I examine the robustness of the classification of "plausibly intended enrollment." In the main text, this category encompasses any municipality for which enrollment falls between the number of individuals in poverty and the population. However, the "plausible" category could also include places with substantial over-enrollment. I examine the robustness of the finding to redefinition of this category. Specifically, I define this category as:

$$\text{Plausible} \in [\text{Poverty Rate}, \min\{j + \text{Poverty Rate}, 1\}] \tag{A.13}$$

for $j \in [0.4, 1]$. Note that the main definition assumes that $j = 1$. The revised scatter plot illustrating this coding is graphed in Figure A.16.

	Complete	Incomplete	Any Info.	Alcaldía Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
PANEL A: CONDITIONAL AMCE BY TERCILE OF POVERTY WITH INTERACTIVE ELECTORAL THREAT CONTROLS					
Lower-Middle Class	0.010	−0.023	−0.013	−0.017	−0.017
	(0.029)	(0.050)	(0.046)	(0.030)	(0.047)
Medium Poverty: Lower-Middle Class	0.026	0.090	0.116*	−0.041	0.081
	(0.044)	(0.074)	(0.068)	(0.044)	(0.064)
High Poverty: Lower-Middle Class	−0.007	0.147**	0.140**	−0.014	−0.028
	(0.044)	(0.070)	(0.068)	(0.040)	(0.061)
Conditional Effect in M	0.036*	0.067*	0.103**	-0.059**	0.064*
	(0.029)	(0.048)	(0.047)	(0.033)	(0.041)
Conditional Effect in H	0.003	0.124***	0.128***	-0.032*	-0.045
	(0.035)	(0.046)	(0.046)	(0.026)	(0.041)
PANEL B: CONDITIONAL AMCE BY TERCILE OF POVERTYWITH INTERACTIVE ELECTORAL THREAT, POPULATION CONTROLS					
Lower-Middle Class	0.015	0.020	0.034	−0.041	0.012
	(0.035)	(0.059)	(0.054)	(0.037)	(0.059)
Medium Poverty: Lower-Middle Class	0.024	0.042	0.066	−0.005	0.048
	(0.050)	(0.083)	(0.080)	(0.054)	(0.075)
High Poverty: Lower-Middle Class	−0.033	0.123	0.089	0.018	−0.060
	(0.052)	(0.087)	(0.087)	(0.057)	(0.087)
Conditional Effect in M	0.039*	0.061	0.101**	-0.046*	0.060*
	(0.031)	(0.050)	(0.048)	(0.036)	(0.041)
Conditional Effect in H	-0.018	0.142***	0.124***	-0.023	-0.048
	(0.036)	(0.050)	(0.050)	(0.034)	(0.048)
Interactive Population Decile Bins	✓	✓	✓	✓	✓
PANEL C: CAMCE BY TERCILE OF POVERTY RATE WITH INTERACTIVE DEPARTMENT FIXED EFFECTS					
Lower-Middle Class	−0.034	0.032	−0.003	−0.035	0.010
	(0.041)	(0.061)	(0.056)	(0.036)	(0.058)
Medium Poverty: Lower-Middle Class	0.072	0.059	0.131	−0.028	0.077
	(0.054)	(0.080)	(0.080)	(0.054)	(0.077)
High Poverty: Lower-Middle Class	0.072	0.097	0.169*	−0.042	−0.006
	(0.066)	(0.098)	(0.091)	(0.062)	(0.084)
Conditional Effect in M	0.038	0.091**	0.129***	-0.063**	0.088**
	(0.033)	(0.051)	(0.052)	(0.038)	(0.045)
Conditional Effect in H	0.038	0.129**	0.166***	-0.077**	0.005
	(0.040)	(0.059)	(0.055)	(0.037)	(0.049)
PANEL D: CAMCE BY TERCILE OF POVERTY RATE WITH INTERACTIVE DEPARTMENT FIXED EFFECTS, POPULATION CONTROLS					
Lower-Middle Class	−0.068	0.160**	0.091	−0.041	0.027
	(0.055)	(0.077)	(0.073)	(0.037)	(0.072)
Medium Poverty: Lower-Middle Class	0.116	−0.077	0.039	−0.005	0.066
	(0.071)	(0.095)	(0.096)	(0.054)	(0.091)
High Poverty: Lower-Middle Class	0.114	−0.105	0.009	0.018	−0.036
	(0.092)	(0.129)	(0.126)	(0.057)	(0.115)
Conditional Effect in M	0.048*	0.083*	0.130***	-0.046*	0.094**
	(0.036)	(0.054)	(0.053)	(0.036)	(0.046)
Conditional Effect in H	0.046	0.055	0.101*	-0.023	-0.008
	(0.046)	(0.071)	(0.069)	(0.034)	(0.061)
Interactive Population Decile Bins	✓	✓	✓	✓	✓
Estimator	IPW	IPW	IPW	IPW	IPW
Observations	1,194	1,194	1,194	1,194	1,194

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.23: Estimates of the conditional AMCE by tercile of poverty, controlling for electoral threats, department, and poverty. The base category is the first tercile (lowest rate of poverty). All covariates and moderators (poverty tercile indicator, demeaned, poverty decile bins, demeaned electoral threat indicators, and demeaned department indicators) are interacted across all factors in the design. All estimates use the IPW estimator. Standard errors are clustered at the municipality level since this is the level of measurement of the poverty moderator.

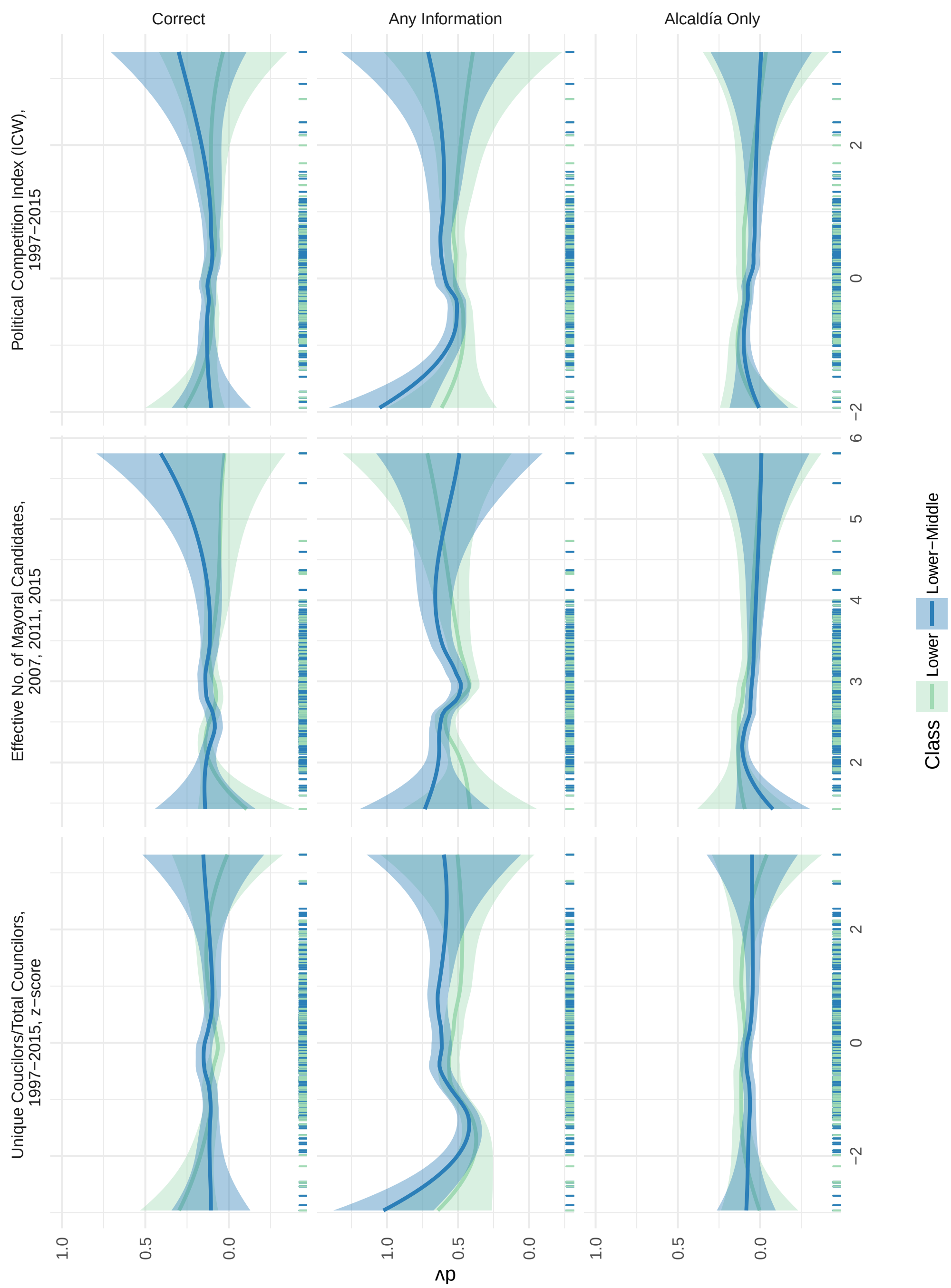


Figure A.15: There is no evidence that class-based bias varies in measures of local political competition. Lines estimated by local polynomial regression (Loess) with 95% confidence intervals.

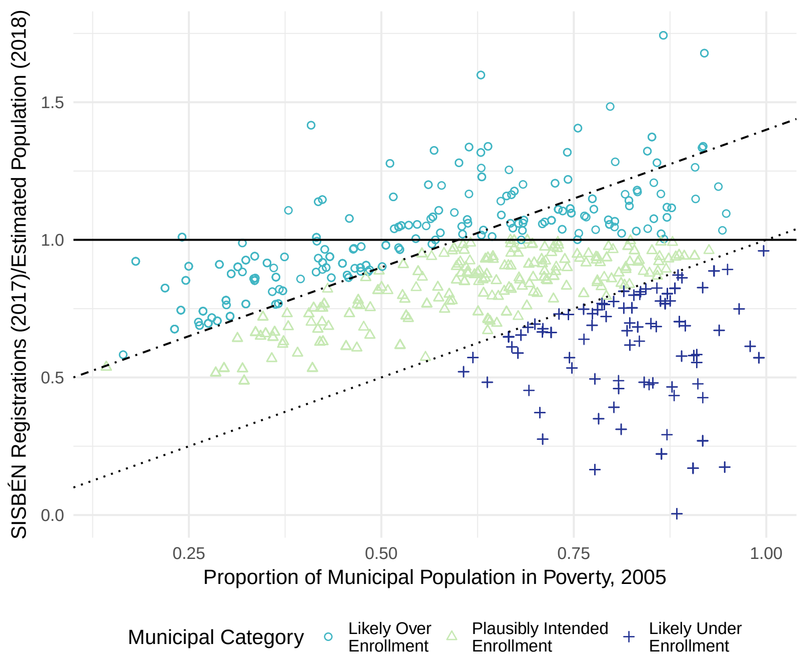


Figure A.16: Visualization of the redefinition of “plausible enrollment” for $j = 0.4$.

Re-estimating Panel A of Table 5 with this specification, Figure A.17 indicates the the point estimates for Plausible Enrollment and the difference (lower panel) are remarkably similar

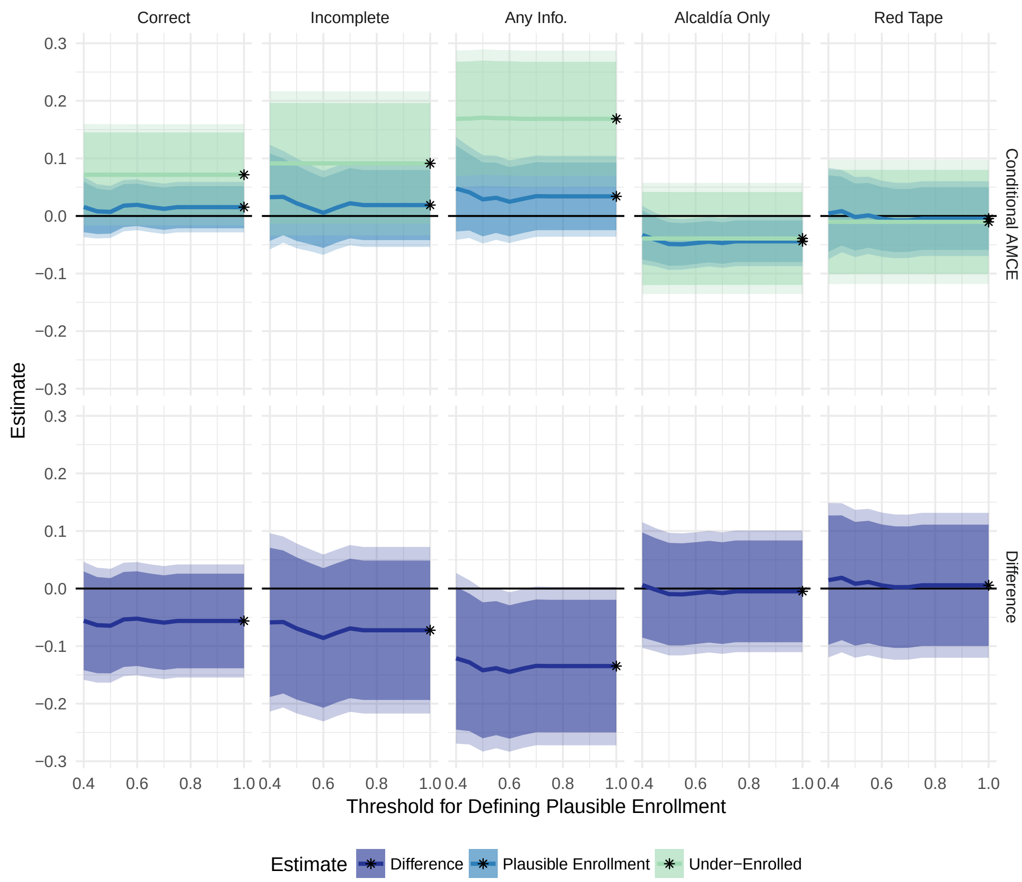


Figure A.17: Robustness of the results in Table 5, Panel A to redefining plausible enrollment." The x-axis corresponds to j in Equation A.13. The stars represent the estimates reported (or implied) by Table 5 Panel A. 90 and 95% confidence intervals calculated on the basis of cluster robust standard errors.

Appendix B

Appendix to Chapter 3

B.1 Proofs

Proof of Proposition 3.1

Consider three cases, defined in terms of partitions of the politician type space, θ_P.

- Case #1: $\theta_P > 1$:

  First, note that a citizen of the politician's type, $\theta = \theta_P > 1$ could never be induced to complain per Equation 3.7. Substituting $\rho(0,0)$ and $\rho(1,0)$ into the politician's objective, the politician maximizes:

$$\max_{\substack{\rho(0,0),\rho(0,1),\\ \rho(1,0),\Delta}} \begin{array}{l} \frac{1}{2}\big[(q+ep)(1-\frac{\rho(1,0)^2}{2}) + (1-q-ep)(\rho(0,0)-\frac{\rho(0,0)^2}{2})+ \\ \quad (q+ep)(1-\frac{\rho(0,0)^2}{2}) + (1-q-ep)(\rho(1,0)-\frac{\rho(1,0)^2}{2})\big] - \frac{\rho(0,1)^2}{2} - \frac{\rho(1,1)^2}{2} \end{array} \tag{B.1}$$

  Optimal monitoring rates $\rho^*(0,0) = (1-q-pe)^2$, $\rho^*(0,1) = 0$, $\rho^*(1,0) = (1-q-pe)$, and $\rho^*(1,1) = 0$ follow from inspection of Equation B.1. Substituting $\rho^*(0,0)$ and $\rho^*(1,0)$, the bureaucrat's IC constraint in Equation 3.9 simplifies to:

$$\Delta \geq \frac{2d}{2p(1-q-p)} = \frac{d}{p(1-q-p)}$$

Note that $\frac{d}{p(1-q-p)} \in (\overline{\Delta}_M, \overline{\Delta}_H]$. As such, effort incentives can only be provided if $\overline{\Delta} = \overline{\Delta}_H$. The bureaucrat's additional IC constraints in Equations 3.13 and 3.15 are satisfied because $\rho^*(0,0) = \rho^*(1,0)$. Define the two resultant contracts (without and with effort incentives) as:

$$\boldsymbol{\varrho}_\emptyset = \{\rho(0,0) = 1-q, \rho(0,1) = 0, \rho(1,0) = 1-q, \rho(1,1) = 0, \overline{\Delta} < \frac{d}{p(1-q-p)}\}$$
$$\boldsymbol{\varrho}_E = \{\rho(0,0) = 1-q-p, \rho(0,1) = 0, \rho(1,0) = 1-q-p, \rho(1,1) = 0, \overline{\Delta} \in [\frac{d}{p(1-q-p)}, \overline{\Delta}_H]\}$$

The difference in the politician's expected utility with and without bureaucratic effort is:

$$E[U_P(\boldsymbol{\varrho}_E)] - E[U_P(\boldsymbol{\varrho}_\emptyset)] = \frac{p(2q+p)}{2} \geq 0$$

Thus, for $\overline{\Delta} = \overline{\Delta}_H$, the politician will implement these monitoring rates of $\rho(0,0) = \rho(1,0) = 1-q-p$ and set $\Delta \in [\frac{d}{p(1-q-p)}, \overline{\Delta}_H]$. For $\overline{\Delta} \in \{\overline{\Delta}_L, \overline{\Delta}_M\}$, the politician will implement these monitoring rates of $\rho(0,0) = \rho(1,0) = 1-q$ and set $\Delta \in [0, \overline{\Delta}]$.

- Case #2: $\theta_P \leq \frac{q+p}{q+p+(1-p-q)^2}$:

First, consider a politician of a type that could always be induced to complain, i.e. $\theta_P \to 0$ if $\omega = 1$ and $a = 0$. Substituting $\rho(0,0)$, $\rho(0,1)$, and $\rho(1,0)$ into the politician's objective, the politician maximizes:

$$\max_{\substack{\rho(0,0),\rho(0,1),\\ \rho(1,0),\Delta}} \begin{aligned} &\frac{1}{2}\big[(q+ep)(1-\frac{\rho(1,0)^2}{2})+(1-q-ep)(\rho(0,1)-\frac{\rho(0,1)^2}{2})+ \\ &(q+ep)(1-\frac{\rho(0,0)^2}{2})+(1-q-ep)(\rho(1,0)-\frac{\rho(1,0)^2}{2})\big]-\frac{\rho(1,1)^2}{2} \end{aligned} \tag{B.2}$$

Maximization of Equation B.2 yields $\rho^*(0,0)=0$, $\rho^*(0,1)=1$, $\rho^*(1,0)=(1-q-ep)$, and $\rho^*(1,1)=0$. However, these monitoring probabilities violates the IC constraint in Equation 3.13:

$$\frac{\rho(1,0)}{\rho(0,1)}=\frac{(1-q-ep)}{1}<\frac{1-q-pe}{q+pe}$$

Violation of this constraint implies that the bureaucrat would not exert effort ($e=0$) and would (i) allocate $a=1$ to all citizens for whom $\theta\leq 1$, per Equation 3.13. Further, the inequality in Equation 3.15 is not satisfied so the bureaucrat allocates $a=0$ to all citizens for whom $\theta>1$ regardless of his investigation.

Per the result in Prendergast (2003), the politician can pursue two alternative contracts. First, consider the case when the politician sets $\Delta=0$ and maintains the optimal monitoring probabilities. Denote this contract $\boldsymbol{\varrho}_I$:

$$\boldsymbol{\varrho}_I=\{\rho(0,0)=0,\rho(0,1)=1,\rho(1,0)=1-q,\rho(1,1)=0,\Delta=0\}$$

In the absence of a penalty, the bureaucrat exerts no effort ($e=0$) and (by assumption) breaks indifference by following his investigation. The politician's expected utility is given by:

$$E[U_P(\boldsymbol{\varrho}_I)]=\frac{2+q+q^2}{4}$$

$E[U_P(\boldsymbol{\varrho}_I)] - E[U_P(\boldsymbol{\varrho}_\emptyset)] = \frac{q-q^2}{4} > 0$ indicating that any politician of type $\theta < 1$ prefers $\boldsymbol{\varrho}_I$ to $\boldsymbol{\varrho}_\emptyset$. Note that any deviation to $\Delta > 0$ induces the to bureaucrat accede to the citizen with certainty granting $a = 1$ per Equation 3.13. This yields an expected utility of $\frac{1+q-q^2}{2} < \frac{2+q+q^2}{4}$. Thus, the politician cannot increase Δ while maintaining optimal monitoring rates.

Alternatively, the politician can provide effort incentives and adjust monitoring rates such that the bureaucrat cannot profitably accede to a prospective complainant. Equation 3.13 provides the relevant IC constraint to ensure that the bureaucrat does not accede to a legible citizen. Maximizing Equation B.2 subject to the constraint implied by Equation 3.13 yields:

$$\rho^*(0,0) = 0, \qquad \rho^*(0,1) = \frac{q+pe}{q+pe+(1-q-pe)^2},$$
$$\rho^*(1,0) = \frac{1-q-pe}{q+pe+(1-q-pe)^2}, \qquad \rho^*(1,1) = 0$$

Substituing $\rho)^*(0,1)$ and $\rho)^*(1,0)$ into the bureaucrat's (other) IC constraint in Equation 3.9 yields:

$$\Delta \geq \frac{2d(q+p+(1-q-p)^2)}{p} = \overline{\Delta}_M$$

Define this contract as:

$$\boldsymbol{\varrho}_{\overline{IE}} = \{\rho(0,0) = 0, \rho(0,1) = \frac{q+p}{q+p+(1-q-p)^2}, \rho(1,0) = \frac{1-q-p}{q+pe+(1-q-pe)^2},$$
$$\Delta \in [\overline{\Delta}_M, \overline{\Delta}]\}$$

The politician's expected utility is:

$$E[U_P(\boldsymbol{\varrho}_{\overline{IE}})] = \frac{1+4p^3+3q-4q^2+4q^3+4p^2(-1+3q)+p(3-8q+12q^2)}{4(q+p+(1-q-p)^2)}$$

The politician cannot profitably deviate by forgoing information as $E[U_P(\boldsymbol{\varrho}_{\overline{IE}})] - E[U_P(\boldsymbol{\varrho}_E)] > 0$. For $\overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\}$, the politician adopts the contract $\boldsymbol{\varrho}_{\overline{IE}}$ if $E[U_P(\boldsymbol{\varrho}_{\overline{IE}})] \geq E[U_P(\boldsymbol{\varrho}_I)]$. Define $\widehat{p}(q)$ as the solution to $E[U_P(\boldsymbol{\varrho}_{\overline{IE}})] = E[U_P(\boldsymbol{\varrho}_I)]$, expressed as a function of q. Note that $\widehat{p}(q) \in [0, \frac{1}{2}]\ \forall q \in [\frac{1}{2}, 1)$ and $\widehat{p}(q) < 1 - q\ \forall q \in [\frac{1}{2}, 1)$.

Note that with incentives, the marginal complainant, is $\theta = \frac{q+p}{p+q+(1-p-q)^2}$. Thus, for any politician of type $\theta_P \leq \frac{q+p}{p+q+(1-p-q)^2}$, the equilibrium contract is given by:

$$\boldsymbol{\varrho} = \begin{cases} \boldsymbol{\varrho}_{\overline{IE}} & \text{if } \overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\} \text{ and } p \geq \widehat{p}(q) \\ \boldsymbol{\varrho}_I & \text{else} \end{cases}$$

- Case #3: $\theta_P \in (\frac{q+p}{q+p+(1-p-q)^2}, 1]$:

Finally, consider politicians of intermediate types, $\theta_P \in (\frac{q+p}{q+p+(1-p-q)^2}, 1]$. Per Equation, 3.7, such a citizen can be induced to complain when $\rho(0,1) - \rho(0,0) \geq \theta_P$. As such, the objective is identical to the previous case. In this case, the contract without incentives ($\boldsymbol{\varrho}_I$) implies that a citizen of type $\theta \in (\frac{q+p}{q+p+(1-p-q)^2}, 1]$ can be induced to complain. This contract follows directly from the proof of the previous case and is thus omitted.

For a contract with effort incentives, however, the politician can only induce complaints by monitoring at higher rates than the contract $\boldsymbol{\varrho}_{\overline{IE}}$. Noting that the optimal $\rho(0,0) = 0$ as above and Equation 3.7, the politician must monitor at the rate $\rho(0,1) = \frac{\theta_P}{y}$ to incentivize complaint from her type of citizen. In this interval, it is straightforward to see that $\frac{\theta_P}{y} > \frac{q+p}{q+p+(1-p-q)^2}$. Substituting $\rho(0,1)$ into Equation 3.13, and rearranging, the politician must set $\rho(1,0) = \frac{\theta_P(1-p-q)}{q+p}$ to satisfy the bureaucrat's "truth-telling" constraint. Substituing $\rho(0,1)^*$ and $\rho(1,0)^*$ into the bureaucrat's (other) IC constraint in Equation 3.9 yields:

$$\Delta \geq \frac{2d(q+p)}{p\theta_P} \in (\overline{\Delta}_L, \overline{\Delta}_M]$$

in the relevant parameter space. Thus, denote the contract with effort incentives:

$$\boldsymbol{\varrho}_{\underline{IE}} = \{\rho(0,0) = 0, \rho(0,1) = \frac{\theta_P}{y}, \rho(1,0) = \frac{\theta_P(1-q-p)}{q+p}, \Delta \in [\overline{\Delta}_M, \overline{\Delta}]\}$$

The politician's expected utility is:

$$\begin{aligned} E[U_P(\boldsymbol{\varrho}_{\underline{IE}})] =& 2\theta_P(q-q^2)(1+\theta_P) - \theta_P^2 + (q^3+p^3)(4y^2+\theta^2) + p^2(-2\theta(y+\theta) + 3q(4y^2+\theta^2)) + \\ & p(2\theta(y+\theta) - 4q\theta(y+\theta) + 3q^2(4y^2+\theta^2)) \end{aligned}$$

As in the previous case, he politician cannot profitably deviate by forgoing information as $E[U_P(\boldsymbol{\varrho}_{\underline{IE}})] - E[U_P(\boldsymbol{\varrho}_E)] > 0$. For $\overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\}$, the politician adopts the contract $\boldsymbol{\varrho}_{\underline{IE}}$ if $E[U_P(\boldsymbol{\varrho}_{\underline{IE}})] \geq E[U_P(\boldsymbol{\varrho}_I)]$. Define $\bar{p}(q)$ as the solution to $E[U_P(\boldsymbol{\varrho}_{\underline{IE}})] = E[U_P(\boldsymbol{\varrho}_I)]$, expressed as a function of q. Note that $\bar{p}(q) \in [0, \frac{1}{2}]\ \forall q \in [\frac{1}{2}, 1)$; $\bar{p}(q) < 1-q\ \forall q \in [\frac{1}{2}, 1)$; and $\bar{p}(q) < \widehat{p}(q)$.

Thus, for any politician of type $\theta_P \in (\frac{q+p}{q+p+(1-p-q)^2}, 1]$, the equilibrium contract is given by:

$$\boldsymbol{\varrho} = \begin{cases} \boldsymbol{\varrho}_{\underline{IE}} & \text{if } \overline{\Delta} \in \{\overline{\Delta}_M, \overline{\Delta}_H\} \text{ and } p \geq \bar{p}(q) \\ \boldsymbol{\varrho}_I & \text{else} \end{cases}$$

■

Proof of Remark 3.1

For $E[C] = 1$, $E[C|\theta] = 1 \forall\theta$. First, consider Contract $\boldsymbol{\varrho}_{\emptyset}$. Under this contract $E[C] = 1 \Rightarrow q + (1-q)^2 = 1$. The unique solution is $q = 1$.

Second, consider contract $\boldsymbol{\varrho}_{E}$. $E[C] = 1 \Rightarrow q + p + (1-q-p)^2 = 1$, which implies $q + p = 1$. Adoption of $\boldsymbol{\varrho}_{E}$ does not occur when $q + p = 1$ because, as $p + q = 1, \overline{\Delta}_H \to \infty$.

Third, consider Contract $\boldsymbol{\varrho}_{E}$. To achieve $E[C] = 1$, it must be the case that (i) $E[C|\theta \leq y] = 1$ and $F(1) = 1$ or (ii) $E[C|\theta \leq y] = E[C|\theta > y] = 1$. (i) requires that $\frac{2-q+q^2}{2} = 1$, which implies that $q \in \{0, 1\}$; only $q = 1$ is within the domain of q and $F(1) = 1$. (ii) implies that $\frac{2-q+q^2}{2} = \frac{(1+q^2)}{2} \Rightarrow q = 1$.

Fourth consider contract $\boldsymbol{\varrho}_{\overline{IE}}$. To achieve $E[C] = 1$ it must be the case that $q+p+\frac{1-p-q}{2(p+q+(1-p-q)^2)} = 1$ and $F(\frac{1-p-q}{2(p+q+(1-p-q)^2)}) = 1$. $q + p + \frac{1-p-q}{2(p+q+(1-p-q)^2)} = 1$ implies that $q + p = 1$.

Finally consider contract $\boldsymbol{\varrho}_{\underline{IE}}$. To achieve $E[C] = 1$ it must be the case that $q + p + \frac{\theta_P(1-p-q)}{q+p} = 1$ and $F(q + p + \frac{\theta_P(1-p-q)}{q+p}) = 1$. $q + p + \frac{\theta_P(1-p-q)}{q+p} = 1$ implies that $q + p = 1$.

■

Proof of Proposition 3.2

Following Table 3.1, the measure of implementation capacity in the presence of information transmission is given by the following expression. With some abuse of notation, the I subscript refers to any equilibrium contract (in a given parameter space) with information transmission.

$$F(\widetilde{\theta})E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}] + (1 - F(\widetilde{\theta}))E[C(\boldsymbol{\varrho}_I)|\theta < \widetilde{\theta}]$$

In the absence of information transmission, implementation capacity is given by:

$$E[C(\boldsymbol{\varrho}_{\neg \boldsymbol{I}})]$$

where the $\neg I$ subscript refers to any equilibrium contract (in a given parameter space) without information transmission.

Denote by λ the share of the legible population, $F(\widetilde{\theta})$, at which implementation capacity is equivalent in contracts with and without information, formally:

$$\lambda(E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}]) + (1 - \lambda)E[C(\boldsymbol{\varrho}_I)|\theta < \widetilde{\theta}] = E[C(\boldsymbol{\varrho}_{\neg I})]$$

A sufficient condition for $\lambda \in [0, 1]$ is $E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}] \geq E[C(\boldsymbol{\varrho}_{\neg I})]$ and $E[C(\boldsymbol{\varrho}_I)|\theta > \widetilde{\theta}] \leq E[C(\boldsymbol{\varrho}_{\neg I})]$.

Proceed by considering cases defined by regions of the parameter space denoted in Proposition 3.1, using the implementation capacity calculations from Table 3.1.

- Case #1: $\overline{\Delta} = \overline{\Delta}_L$

In this case, compare implementation capacity under Contract $\boldsymbol{\varrho}_I$ to Contract $\boldsymbol{\varrho}_\emptyset$:

$$E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}] = \frac{2-q+q^2}{2}$$
$$E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}] = \frac{1+q^2}{2}$$
$$E[C(\boldsymbol{\varrho}_\emptyset)] = 1-q+q^2$$

Clearly, for any $q \in [\frac{1}{2}, 1)$, $\frac{1+q^2}{2} < 1-q+q^2 < \frac{2-q+q^2}{2}$, which is sufficent for $\lambda \in [0,1]$.

- Case #2: $\overline{\Delta} = \overline{\Delta}_M$ and $p < \widehat{p}(q)$:

  This case is identical to the previous case for any θ_P and is thus omitted.

- Case #3: $\overline{\Delta} = \overline{\Delta}_M$ and $p \in [\widehat{p}(q), \overline{p}(q))$:

  In this case, compare implementation capacity under Contract $\boldsymbol{\varrho}_{\overline{IE}}$ to $\boldsymbol{\varrho}_\emptyset$:

$$E[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta \leq \widetilde{\theta}] = q+p+\frac{1-p-q}{2(q+p+(1-p-q)^2)}$$
$$E[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta > \widetilde{\theta}] = \frac{1}{2}$$
$$E[C(\boldsymbol{\varrho}_\emptyset)] = 1-q+q^2$$

  For any $q \in [\frac{1}{2}, 1)$, $p \in (\widehat{p}(q), 1-q]$, $\frac{1}{2} < 1-q+q^2 < q+p+\frac{1-p-q}{2(q+p+(1-p-q)^2)}$. The latter inequality holds when $p=0$ and note that $\frac{dE[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta \leq \widetilde{\theta}]}{dp} > 0$.

  The comparison between Contract $\boldsymbol{\varrho}_I$ and Contract $\boldsymbol{\varrho}_\emptyset$ is equivalent to Case #1 and is therefore omitted.

- Case #4: $\overline{\Delta} = \overline{\Delta}_M$ and $p \geq \overline{p}(q)$:

  The analysis of $\boldsymbol{\varrho}_{\overline{IE}}$ and $\boldsymbol{\varrho}_\emptyset$ is identical to the previous case.

Compare implementation capacity under Contract $\boldsymbol{\varrho}_{\underline{IE}}$ to Contract $\boldsymbol{\varrho}_{\emptyset}$:

$$E[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta \leq \widetilde{\theta}] = q + p + \frac{\theta_P(1-p-q)}{2(q+p)}$$
$$E[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta > \widetilde{\theta}] = \frac{1}{2}$$
$$E[C(\boldsymbol{\varrho}_{\emptyset})] = 1 - q + q^2$$

For any $q \in \frac{1}{2}, 1)$, $p \in (\widehat{p}(q), 1-q]$, $\frac{1}{2} < 1 - q + q^2 < q + p + \frac{\theta_P(1-p-q)}{2(q+p)}$. By Propositoin 3.1, in this parameter region, $\frac{\theta_P(1-p-q)}{2(q+p)} > q + p + \frac{1-p-q}{2(q+p+(1-p-q)^2)}$ from the previous case. Combined with the previous case, this is sufficient for the the inquality to hold.

- Case #5: $\overline{\Delta} = \overline{\Delta}_H, p < \widehat{p}(q)$: Compare implementation under Contract $\boldsymbol{\varrho}_I$ to Contract $\boldsymbol{\varrho}_E$ as follows:

$$E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}] = \frac{2 - q + q^2}{2}$$
$$E[C(\boldsymbol{\varrho}_I)|\theta \leq \widetilde{\theta}] = \frac{1 + q^2}{2}$$
$$E[C(\boldsymbol{\varrho}_{\emptyset})] = q + p + (1 - p - q)^2$$

For any $q \in \frac{1}{2}, 1)$, $p \in (\widehat{p}(q), 1-q]$, $\frac{1+q^2}{2} < q + p + (1-p-q)^2$. If $p < \frac{1}{2}(1 - 2q + \sqrt{1 - 2q + 2q^2})$, $\frac{2-q+q^2}{2} > q + p + (1-p-q)^2$. This condition is satsfied for any $p < \widehat{p}(q)$.

- Case #6: $\overline{\Delta} = \overline{\Delta}_H, p \in [\widehat{p}(q), \overline{p}(q))$:

The comparison of Contracts $\boldsymbol{\varrho}_I$ to $\boldsymbol{\varrho}_E$ is identical to the previous case, though note that $p < \frac{1}{2}(1 - 2q + \sqrt{1 - 2q + 2q^2})$ is also satisfied for any $p < \overline{p}(q)$.

Compare implementation capacity under contracts $\boldsymbol{\varrho}_{\overline{IE}}$ and $\boldsymbol{\varrho}_E$:

$$E[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta \leq \widetilde{\theta}] = q + p + \frac{1-p-q}{2(q+p+(1-p-q)^2)}$$
$$E[C(\boldsymbol{\varrho}_{\overline{IE}})|\theta > \widetilde{\theta}] = \frac{1}{2}$$
$$E[C(\boldsymbol{\varrho}_E)] = q + p + (1-q-p)^2$$

For any $q \in \frac{1}{2}, 1)$, $p \in (\widehat{p}(q), 1-q]$, it is clear from inspection that $\frac{1}{2} < q+p+(1-q-p)^2 < q+p+\frac{1-p-q}{2(q+p+(1-p-q)^2)}$.

- Case #7: $\overline{\Delta} = \overline{\Delta}_H, p \geq \overline{p}(q)$: The comparison of implementation capacity under Contracts $\boldsymbol{\varrho}_{\overline{IE}}$ and $\boldsymbol{\varrho}_E$ is equivalent to the previous case.

  Compare implementation capacity under contracts $\boldsymbol{\varrho}_{\underline{IE}}$ and $\boldsymbol{\varrho}_E$:

$$E[C(\boldsymbol{\varrho}_{\underline{IE}})|\theta \leq \widetilde{\theta}] = q + p + \frac{\theta_P(1-p-q)}{2(q+p)}$$
$$E[C(\boldsymbol{\varrho}_{\underline{IE}})|\theta > \widetilde{\theta}] = \frac{1}{2}$$
$$E[C(\boldsymbol{\varrho}_E)] = q + p + (1-q-p)^2$$

  For any $q \in \frac{1}{2}, 1)$, $p \in (\widehat{p}(q), 1-q]$, it is clear from inspection that $\frac{1}{2} < q+p+(1-q-p)^2 < q+p+\frac{\theta_P(1-p-q)}{2(q+p)}$.

■

Proof of Proposition 3.3

First, note that the area of the triangle defined by the coordinates in Definition 3.2 is given by:

$$\mu_2\left((0,0), (F(\widetilde{\theta}), \frac{F(\widetilde{\theta})E[a^\dagger|\theta \leq \widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta \leq \widetilde{\theta}] + (1-F(\widetilde{\theta}))E[a^\dagger|\theta > \widetilde{\theta}]}), (1,1)\right)$$
$$= \frac{1}{2}\left(\frac{F(\widetilde{\theta})E[a^\dagger|\theta \leq \widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta \leq \widetilde{\theta}] + (1-F(\widetilde{\theta}))E[a^\dagger|\theta > \widetilde{\theta}]} - F(\widetilde{\theta})\right)$$

Consider each of the five contracts. For Contracts $\boldsymbol{\varrho}_{\emptyset}$ and $\boldsymbol{\varrho}_E$, $E[a^\dagger|\theta]$ is equivalent for all a. This implies that the point $F(\widetilde{\theta}) = \frac{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]+(1-F(\widetilde{\theta}))E[a^\dagger|\theta>\widetilde{\theta}]}$. Thus, $TAI(\boldsymbol{\varrho}_1) = TAI(\boldsymbol{\varrho}_2) = 0$ and $\frac{\partial TAI(\boldsymbol{\varrho}_1)}{\partial q} = \frac{\partial TAI(\boldsymbol{\varrho}_2)}{\partial q} = 0$.

Under Contract $\boldsymbol{\varrho}_I$:

$$\begin{aligned}\frac{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]+(1-F(\widetilde{\theta}))E[a^\dagger|\theta>\widetilde{\theta}]} &= \frac{F(1)(2-q+q^2)}{F(1)-F(1)q+(2-q)q}\\ \Rightarrow TAI(\boldsymbol{\varrho}_I) &= \frac{F(1)(2-q+q^2)}{F(1)-F(1)q+(2-q)q} - F(1)\\ &= \frac{F(1)(1-F(1))(1-q)}{F(1)(1-q)+(2-q)q} > 0\end{aligned}$$

Note that $\frac{\partial TAI(\boldsymbol{\varrho}_I)}{\partial q} = \frac{(F(1)-1)F(1)(2-2q+q^2)}{(F(1)(q-1)+(q-2)q)^2} < 0$.

Under Contract $\boldsymbol{\varrho}_{\overline{IE}}$, $E[a^\dagger|\theta < \frac{q+p}{q+p+(1-q-p)^2}] = 0$. Therefore:

$$\begin{aligned}\frac{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]+(1-F(\widetilde{\theta}))E[a^\dagger|\theta>\widetilde{\theta}]} &= 1\\ \Rightarrow TAI(\boldsymbol{\varrho}_{\overline{IE}}) &= 1 - F\left(\frac{q+p}{q+p+(1-q-p)^2}\right)\end{aligned}$$

Note that $\frac{\partial TAI(\boldsymbol{\varrho}_{\overline{IE}})}{\partial q} = -f(\frac{q+p}{q+p+(1-q-p)^2})\frac{1-p^2-q^2-2qp}{(p+q+(1-p-q)^2)^2}$. $f(\cdot)$ is the pdf of θ and is non-negative. As such $\frac{\partial TAI(\boldsymbol{\varrho}_4)}{\partial q} \leq 0$.

Finally, under Contract $\boldsymbol{\varrho}_{\underline{IE}}$, $E[a^\dagger|\theta < \frac{\theta_P}{y}] = 0$. Therefore:

$$\begin{aligned}\frac{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]}{F(\widetilde{\theta})E[a^\dagger|\theta\leq\widetilde{\theta}]+(1-F(\widetilde{\theta}))E[a^\dagger|\theta>\widetilde{\theta}]} &= 1\\ \Rightarrow TAI(\boldsymbol{\varrho}_{\underline{IE}}) &= 1 - F\left(\frac{\theta_P}{y}\right)\end{aligned}$$

Note that $\frac{\partial TAI(\boldsymbol{\varrho}_{\underline{IE}})}{\partial q} = 0$.

Comparing TAI across the contracts with information transmission, $TAI(\boldsymbol{\varrho}_{\overline{IE}}) > TAI(\boldsymbol{\varrho}_I)$ can be seen by inspection of the inequality:

$$1 - F\left(\frac{q+p}{q+p+(1-q-p)^2}\right) > \frac{F(1)(2-q+q^2)}{F(1)-F(1)q+(2-q)q} - F(1)$$

Note that $1 > \frac{F(1)(2-q+q^2)}{F(1)-F(1)q+(2-q)q}$ and $F\left(\frac{q+p}{q+p+(1-q-p)^2}\right) \leq F(1)$. Similarly, $TAI(\boldsymbol{\varrho}_{\underline{IE}}) > TAI(\boldsymbol{\varrho}_I)$ is given by:

$$1 - F\left(\frac{\theta_P}{y}\right) > \frac{F(1)(2-q+q^2)}{F(1)-F(1)q+(2-q)q} - F(1)$$

as $1 > \frac{F(1)(2-q+q^2)}{F(1)-F(1)q+(2-q)q}$ and $F\left(\frac{\theta_P}{y}\right) \leq F(1)$.

■

Proof of Proposition 3.4

Optimal contracts and incentives, by citizen type, follow directly from Proposition 3.1 and are represented in Table B.1.

Contract	Parameter region $\overline{\Delta}$	p	Citizen type, θ: $\theta \leq \frac{q+p}{p+q+(1-p-q)^2}$	$\theta \in (\frac{q+p}{p+q+(1-p-q)^2}, 1]$	$\theta > y$
ς_1	$\overline{\Delta}_L$	any	$\boldsymbol{\varrho}_I$	$\boldsymbol{\varrho}_I$	$\boldsymbol{\varrho}_\emptyset$
	$\overline{\Delta}_M$	$p < \widehat{p}(q)$			
ς_2	$\overline{\Delta}_M$	$p \in [\widehat{p}(q), \overline{p}(q))$	$\boldsymbol{\varrho}_{\overline{IE}}$	$\boldsymbol{\varrho}_I$	$\boldsymbol{\varrho}_\emptyset$
ς_3	$\overline{\Delta}_M$	$p > \overline{p}(q)$	$\boldsymbol{\varrho}_{\overline{IE}}$	$\boldsymbol{\varrho}_{\underline{IE}}$	$\boldsymbol{\varrho}_\emptyset$
ς_4	$\overline{\Delta}_H$	$p < \widehat{p}$	$\boldsymbol{\varrho}_I$	$\boldsymbol{\varrho}_I$	$\boldsymbol{\varrho}_E$
ς_5	$\overline{\Delta}_H$	$p \in [\widehat{p}, \overline{p})$	$\boldsymbol{\varrho}_{\overline{IE}}$	$\boldsymbol{\varrho}_I$	$\boldsymbol{\varrho}_E$
ς_6	$\overline{\Delta}_H$	$p \geq \overline{p}$	$\boldsymbol{\varrho}_{\overline{IE}}$	$\boldsymbol{\varrho}_{\underline{IE}}$	$\boldsymbol{\varrho}_E$

Table B.1: Optimal contracts, by citizen type.

Each contract imposes monitoring rates of $\rho(0,1) > \rho(0,0)$ for some citizen type by employing contracts $\boldsymbol{\varrho}_I$, $\boldsymbol{\varrho}_{\overline{IE}}$ or $\boldsymbol{\varrho}_{\underline{IE}}$. Given the assumption $F(1) \in (0,1)$, $\rho(0,1) > \rho(0,0)$ induces some citizen to complain.

Per Definition 3.2, a sufficient condition for $TAI > 0$ is that $\exists \theta', \theta'' \in supp(f)$ such that $E[a^\dagger|\theta'] \neq E[a^\dagger|\theta'']$. The expressions for $E[a^\dagger|\theta]$ in 3.1, indicate for any $F(1) \in (0,1)$ and $q < 1$, $TAI > 0$ for contracts $\boldsymbol{\varrho}_1$ to $\boldsymbol{\varrho}_6$.

Finally, compare the levels of inequality generated by the contracts in Table B.1 to inequality generated by their any constituent contract with information. For contract with two "constituent" contracts (ς_1 and ς_4), these comparisons are straightforward:

$$TAI(\boldsymbol{\varsigma}_1) - TAI((\boldsymbol{\varrho}_I) = \frac{F(1)(1+q-q^2)}{1+F(1)(q-q^2)} - F(1) - \left[\frac{F(1)(1+q-q^2)}{(2-q)q+F(1)(1-q)} - F(1)\right] < 0$$
$$TAI(\boldsymbol{\varsigma_2}) - TAI(\boldsymbol{\varrho}_I)) = \frac{F(1)(1+q-q^2)}{1+F(1)(q-q^2)} - F(1) - \left[\frac{F(1)(1+q-q^2)}{(2-q)q+F(1)(1-q)} - F(1)\right] < 0$$

In the case of contracts $\boldsymbol{\varsigma}_2$, $\boldsymbol{\varsigma}_3$, $\boldsymbol{\varsigma_5}$, and $\boldsymbol{\varsigma}_6$ there exist two thresholds defining different contracts that are applied across the population. Geometrically, the area measure relevant to the calculation of TAI is depicted for the quadrilateral representing $\boldsymbol{\varsigma}_3$ in Figure B.1.

Given the notation in the Figure, $TAI(\boldsymbol{\varrho})$ is equivalent to:

$$2\mu_2\left((0,0),(F(\widetilde{\theta}_1),{a_1}^\dagger),(F(\widetilde{\theta}_2),{a_2}^\dagger),(1,1)\right) = {a_2}^\dagger(1-F(\widetilde{\theta}_1)) + F(\widetilde{\theta}_2)({a_1}^\dagger - 1)$$

Note two observations about contracts $\boldsymbol{\varsigma}_2$, $\boldsymbol{\varsigma}_3$, $\boldsymbol{\varsigma_5}$, and $\boldsymbol{\varsigma}_6$. Each includes contract $\boldsymbol{\varrho}_{\overline{IE}}$ for some

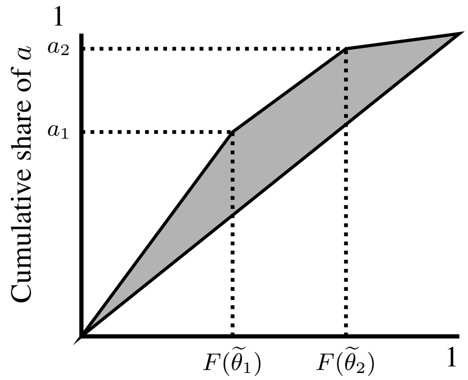


Figure B.1: Geometric representation of inequality measure with three partitions of the type space (x-axis). Note that TAI is equivalent to double the shaded area.

segment of the population. In this case, $F(\tilde{\theta}_1)$ is equivalent as the marginal complainant under contract $\boldsymbol{\varrho}_{\overline{IE}}$ is the same. As such, TAI is greater under $\boldsymbol{\varrho}_{\overline{IE}}$ than under any contract in: ς_2, ς_3, ς_5, ς_6 if:

$$1 - F(\widetilde{\theta}_1) > a_2^{\dagger}(1 - F(\widetilde{\theta}_1)) + F(\widetilde{\theta}_2)(a_1^{\dagger} - 1)$$

This inequality is always satisfied since $a_2^{\dagger} < 1$ and $a_1^{\dagger} < 1$.

■

B.2 Empirical Motivation, Continued

The text provides three possible reasons for variation in complaint rates across geographic units in the same city. I will refer to these units as "neighborhoods."

1. Different rates of service utilization as a function of characteristics of the composition of neighborhood residents. Some neighborhoods' populations may be more reliant on specific government services (or types of government services) than others.

2. Different quality of service provision across different neighborhoods. Lower quality services may yield more complaints.

3. Citizens vary in their costs of complaint in a manner that correlates with the neighborhoods they live in.

To assess these possibilities, I descriptively examine rates of complaint across geographic units in Bogotá and New York, by wealth of residents. I use this measure because wealth is believed to correlate positively with service provision (#2) in both cities. The relative magnitude of this correlation across the cities, however, is not evident.

To examine #1, I disaggregate complaints by the agency to which complaints were directed. This is provided in both datasets. I plot this data descriptively in Figure B.2. Clear gradients emerge in the usage of some services as a function of neighborhood (unit) wealth. In particular, departments of housing are the most frequent recipient of complaints in poorer neighborhoods. This makes sense as their services are disproportionately used by lower-income citizens in both cities. This supports the argument in #1 but it does not speak to variation in rates of complaint.

Figure B.3 looks at variation in the volume of complaint by neighborhood wealth with and without housing-related complaints. In Bogotá, despite widespread evidence and perceptions that services are *better* in rich localities, they also file substantially more complaints than poor localities. The relationship looks quite similar with and without housing complaints given the low rate of complaints from poor localities in general.

In New York, the relationship between census tract wealth and complaint-filing is less clear. Overall, there appears to be slight non-monotonicity in complaint rates by neighborhood wealth. This occurs even though popular wisdom holds that service quality is *increasing* in neighborhood wealth. When health complaints are omitted, a positive correlation between neighborhood wealth and complaint rate emerges, albeit at a lower magnitude than in Bogotá.

The inverse relationship between service quality and complaint rates in both cities, even when adjusting (symmetrically) for different types of service utilization, suggests variation in the propensity of citizens to make complaints. The model captures these tendencies in terms of costs of complaint.

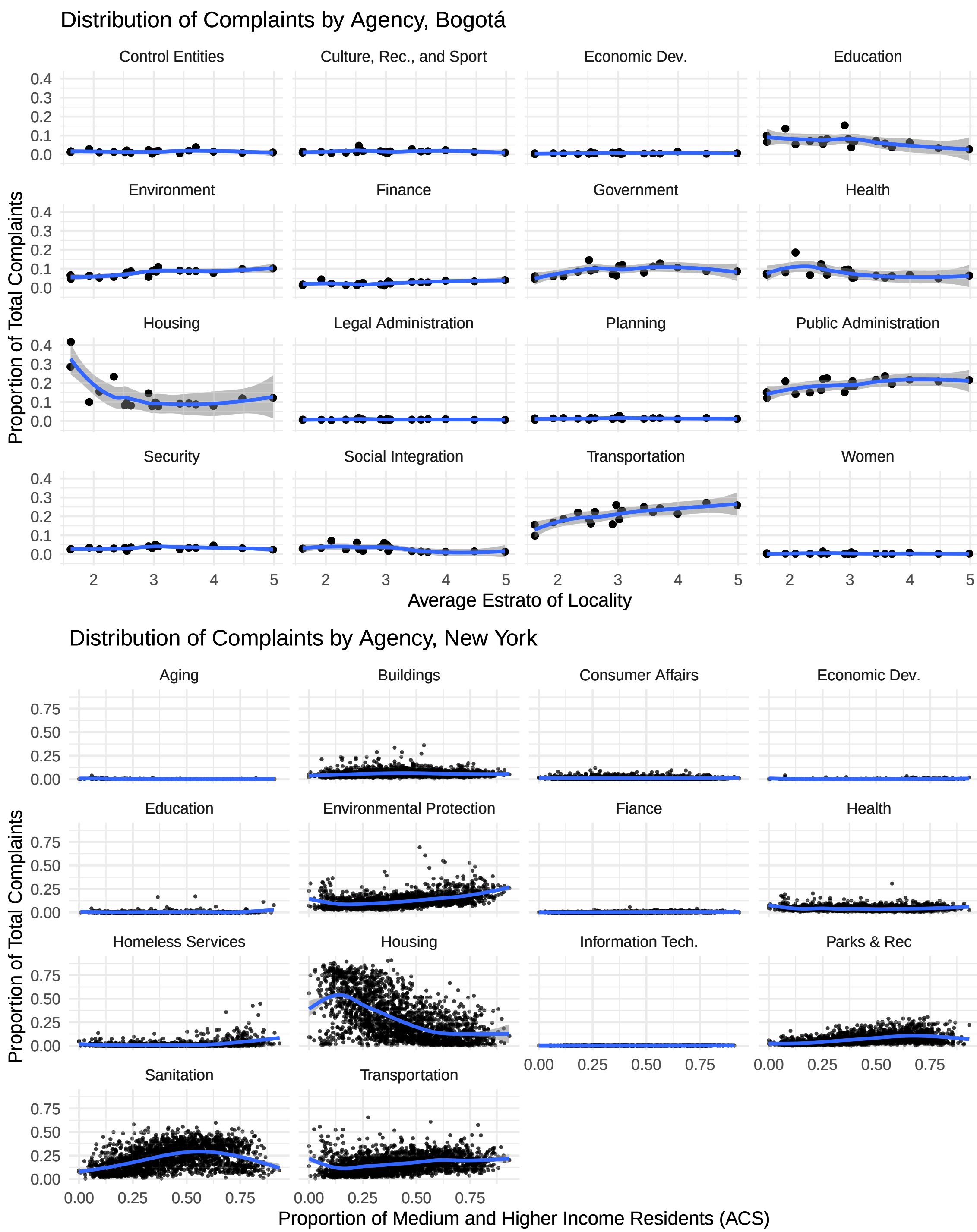


Figure B.2: Proportion of total complaints directed to each city government agency in Bogotá and New York, by neighborhood wealth. The x-axis is increasing in neighborhood wealth. The complaints are aggregated over the January 2017-June 2018 period in both cities.

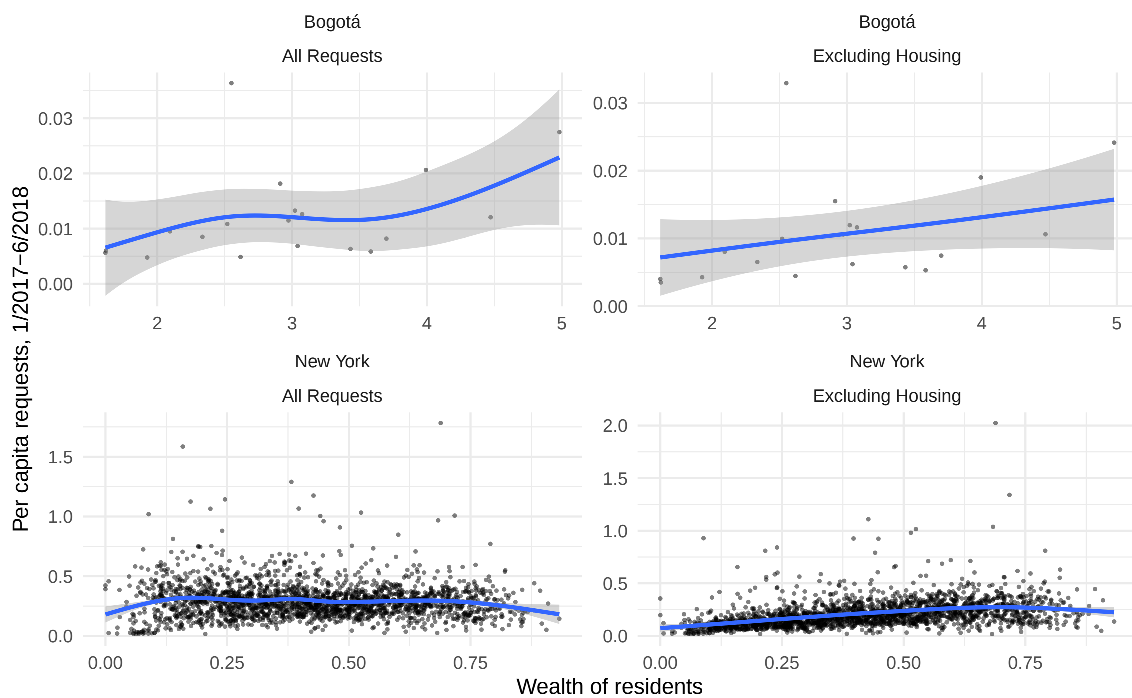


Figure B.3: Per-capita rate of complaint-making by neighborhood wealth. The x-axis is increasing in neighborhood wealth. The rows index the two cities and the columns report the rate of "all complaints" and all non-housing complaints.

Appendix C

Appendix to Chapter 4

C.1 Proofs

First, note that Lemma C.1 is useful in the proofs of Propositions 4.1 and 4.3.

Lemma C.1. *The incumbent's probability of victory,* $\frac{\tau(\mu,\boldsymbol{a})}{\mu}$ *is weakly increasing in the voter's posterior belief,* μ, $\frac{\partial\tau(\mu,\boldsymbol{a})}{\partial\mu} \geq 0$.

Proof: Differentiating Equation 4.9 with respect to μ yields:

$$\frac{\partial\tau(\mu,\mathbf{a})}{\partial\mu} = \frac{E[g_2^i|\theta=\overline{\theta}] - E[g_2^i|\theta=\underline{\theta}]}{2b}$$

$E[g_2^i|\theta = \overline{\theta}] \geq E[g_2^i|\theta = \underline{\theta}]$ follows from the parametric assumption $\overline{\theta} > \underline{\theta}$ and Equation 4.7. Therefore, $\frac{\partial\tau(\mu,\mathbf{a})}{\partial\mu} \geq 0$. ■

Proof of Proposition 4.1

This proof proceeds in two sections. I first prove the existence of the equilibria characterized in Proposition 4.1, then I prove uniqueness.

Existence

First, suppose that $q < \frac{1}{\theta}$ and consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter re-elects the incumbent if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe g_1, $\mu = \pi$;
- Upon observation that $g_1 = 0$, $\mu(g_t = 0) = \pi$;
- Off the equilibrium path, an observation that $g_1 = q$ implies that $\mu(g_t = q) = 1$.

By inspection, μ is derived via Bayes' rule. The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. Given the voter's posterior belief $\mu = \pi$, $\tau(\pi, \mathbf{a}) = \frac{1}{2}$. $q < \frac{1}{\overline{\theta}}$ implies that $E[g_1] < a_1 \forall \theta$, such that a politician of either type allocates $a_1 = 0$. The competent type cannot profitably deviate by allocating $a_1 = 1$ because:

$$1 + \tau(\pi, \mathbf{a}) \geq \overline{\theta} q + p\tau(1, \mathbf{a}) + (1 - p)\tau(\pi, \mathbf{a})$$

In this region where $q < \frac{1}{\overline{\theta}}$, $\overline{\theta} q < 1$ and $\tau(\mu, \mathbf{a}) = \frac{1}{2} \forall \mu$ when $a_1 = a_2 = 0 \forall \theta$. Since $\overline{\theta} > \underline{\theta}$, it therefore holds that the incompetent type similarly cannot profitably deviate by allocating $a_1 = 1$.

Second, suppose that $q \in \left[\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta}) + p\overline{\theta}(1-\pi))}\right)$ and consider the following strategy and belief profile: a politician of type $\theta = \overline{\theta}$ allocates $a_1 = a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates $a_1 = a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe g_1, $\mu = \pi$;
- Upon observation that $g_1 = q$, $\mu(g_1 = q) = 1$;
- Upon observation that $g_1 = 0$, $\mu(g_1 = 0) = \frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + 1 - \pi}$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**, a politician of type $\theta = \overline{\theta}$ will not deviate from $a_1 = 1$ to $a_1 = 0$ since:

$$\overline{\theta}q + \left(p\overline{\theta}\tau(1, \mathbf{a}) + p(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})\right)\overline{\theta}q >$$
$$1 + \left(p\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})\right)\overline{\theta}q.$$

This condition clearly obtains for any $q \in \left[\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\right)$ since $\overline{\theta}q > 1$ and, by Lemma C.1, $\tau(1, \mathbf{a}) > \tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi}, \mathbf{a})$. A politician of type $\theta = \underline{\theta}$ cannot profitably deviate to allocate $a_1 = 1$ to increase her chances of re-election when:

$$1 + p\tau\left(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi}, \mathbf{a}\right) + (1-p)\tau(\pi, \mathbf{a}) >$$
$$\underline{\theta}q + p\underline{\theta}\tau(1, \mathbf{a}) + p(1-\underline{\theta})\tau\left(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi}, \mathbf{a}\right) + (1-p)\tau(\pi, \mathbf{a})$$
$$\Leftrightarrow q < \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}$$

Third, suppose that $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right)$ and consider the following strategy and belief profile: a politician of type $\theta = \overline{\theta}$ allocates $a_1 = a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates $a_1 = 1$ and $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe g_1, $\mu = \pi$;
- Upon observation that $g_1 = q$, $\mu(g_1 = q) = \frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}}$;
- Upon observation that $g_1 = 0$, $\mu(g_1 = 0) = \frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})}$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**, a politician of type $\theta = \overline{\theta}$ will not deviate from $a_1 = 1$ to $a_1 = 0$ since:

$$\overline{\theta}q + \left(p\overline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta} + (1-\pi)\underline{\theta}}, \mathbf{a}) + p(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})\right)\overline{\theta}q >$$
$$1 + \left(p\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})\right)\overline{\theta}q.$$

This inequality holds for any $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right)$ given that $\overline{\theta}q > 1$ and, by Lemma C.1, $\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}}, \mathbf{a}) > \tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})}, \mathbf{a})$. A politician of type $\theta = \underline{\theta}$ cannot profitably deviate by allocating $a_1 = 0$ if:

$$\underline{\theta}q + p\underline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\underline{\theta} + (1-\pi)\underline{\theta}}, \mathbf{a}) + p(1-\underline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}) >$$
$$1 + p\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})$$
$$\Rightarrow q + p\left(\tau(\frac{\pi\overline{\theta}}{\pi\underline{\theta} + (1-\pi)\underline{\theta}}, \mathbf{a}) - \tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a})\right) > \frac{1}{\underline{\theta}}$$
$$q > \frac{2b((\overline{\theta}-\underline{\theta})\pi + \underline{\theta})(-1 + (\overline{\theta}-\underline{\theta}) + \underline{\theta})}{\underline{\theta}[2b((\overline{\theta}-\underline{\theta})\pi + \underline{\theta})(-1 + (\overline{\theta}-\underline{\theta}) + \underline{\theta})) + (\overline{\theta}-\underline{\theta})\overline{\theta}p(1-\pi)\pi]}$$

This inequality holds for any $q \geq \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}$, which is sufficient to ensure that the inequality is satisfied for any $q \in [\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}})$.

Finally, suppose that $q \geq \frac{1}{\underline{\theta}}$ and consider the following strategy and belief profile: politicians of both types allocate $a_1 = a_2 = 1$; the bureaucrat exerts effort proportional to θ in each period; this yields no public goods $g_t = 0 \forall t$; the voter votes to re-elect if $E[u_2^V(i)] > E[u_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe g_1, $\mu = \pi$;
- Upon observation that $g_1 = q$, $\mu(g_t = q) = \frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}}$;
- Upon observation that $g_1 = 0$, $\mu(g_t = 0) = \frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})}$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**, a politician of type $\theta = \underline{\theta}$ will not deviate from $a_1 = 1$ to $a_1 = 0$ since:

$$\underline{\theta}q + p\underline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\underline{\theta} + (1-\pi)\underline{\theta}}, \mathbf{a}) + p(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}) >$$
$$1 + p\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta}) + (1-\pi)(1-\underline{\theta})}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})$$

This inequality holds for any $q \geq \frac{1}{\underline{\theta}}$ because $\underline{\theta}q > 1$ and, by Lemma C.1, $\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}}, \mathbf{a}) > \tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})}, \mathbf{a})$. This is condition, combined with the parametric assumption that $\overline{\theta} > \underline{\theta}$, is sufficient to ensure that a politician of type $\theta = \overline{\theta}$ similarly does not deviate.

Uniqueness Consider first the candidate pooling equilibria and then the possible separating equilibria. First, note that $a_1 = 1 \Rightarrow g_1 \in \{0, q\}$ and $a_1 = 0 \Rightarrow g_1 = 0$. This implies that off-path beliefs are only invoked in an equilibrium in which both types allocates $a_1 = 0$. Per the intuitive criterion refinement, I impose the off-path belief that $\mu = 1$ upon observation that $g_1 = q$ in any such equilibrium. There are three possible equilibria in which both types allocate $a_1 = 0$ that emerge in different regions of the parameter space, depending on a politician's second-period allocation behavior. The first is an equilibrum (the first case of Proposition 4.1), the others are not, as shown below:

- First, suppose $q \in \left[\frac{1}{\overline{\theta}}, \frac{1}{\underline{\theta}}\right)$.

  Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and a politician of type $\theta = \overline{\theta} = 1$ allocates $a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

  - If the voter does not observe g_1, $\mu = \pi$;
  - Upon observation that $g_1 = 0$, $\mu(0) = \pi$;

– Off the equilibrium path, an observation that $g_1 = q$ implies that $\mu = 1$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, $\mathbf{a}$. A politician of type $\theta = \overline{\theta}$ will not deviate if:

$$1 + \tau(\pi, \mathbf{a}) \geq \overline{\theta}q + p\overline{\theta}\tau(1, \boldsymbol{a}) + p(1-\overline{\theta})\tau(\pi, \boldsymbol{a}) + (1-p)\tau(\pi, \boldsymbol{a})$$

However, this inequality is never satisfied since $\overline{\theta}q \geq 1$ and $\tau(1, \mathbf{a}) > \tau(\pi, \mathbf{a})$. Thus, this strategy and belief profile is not an equilibrium.

- Second, suppose $q \geq \frac{1}{\underline{\theta}}$.

  Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and a politician of both types allocate $a_2 = 1$. All other beliefs and strategies are equivalent to the previous case.

  The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, $\mathbf{a}$. A politician of type $\theta = \overline{\theta}$ will not deviate if:

$$1 + \tau(\pi, \mathbf{a}) \geq \overline{\theta}q + p\overline{\theta}\tau(1, \boldsymbol{a}) + p(1-\overline{\theta})\tau(\pi, \boldsymbol{a}) + (1-p)\tau(\pi, \boldsymbol{a})$$

  However, this inequality is never satisfied since $\overline{\theta}q \geq 1$ and $\tau(1, \mathbf{a}) > \tau(\pi, \mathbf{a})$. Thus, this strategy and belief profile is not an equilibrium.

Now consider possible candidate pooling equilibria in which both types allocate $a_1 = 1$. There are three possible equilibria in which both types allocate $a_1 = 0$ that emerge in different regions of the parameter space, depending on a politician's second-period allocation behavior. The third is the final case of Equilibrium 4.1.

- First, suppose $q < \frac{1}{\overline{\theta}}$.

  Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 2$ and either type of politician allocates $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

  - If the voter does not observe g_1, $\mu = \pi$;
  - Upon observation that $g_1 = q$, $\mu(g_1 = q) = \frac{\overline{\theta}\pi}{\overline{\theta}\pi + (1-\pi)\underline{\theta}}$;
  - Upon observation that $g_1 = 0$, $\mu(g_1 = 0) = \frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi + (1-\underline{\theta})(1-\pi)}$.

  The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \underline{\theta}$ will not deviate if:

$$\underline{\theta}q + p\underline{\theta}\tau(\frac{\overline{\theta}\pi}{\overline{\theta}\pi + (1-\pi)\underline{\theta}}, \mathbf{a}) + p(1-\underline{\theta})\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi + (1-\underline{\theta})(1-\pi)}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}) \geq$$
$$1 + p\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi + (1-\underline{\theta})(1-\pi)}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})$$

  In any equilibrium in which $a_2 = 0 \forall\theta$, $\tau(\mu, \boldsymbol{a})$ is equivalent for any μ, per Equation 4.9. Combined $\underline{\theta}q < 1$ in this parameter space, this inequality never holds. Thus, this strategy and belief profile is not an equilibrium.

- Second, suppose $q \in [\frac{1}{\overline{\theta}}, \frac{1}{\underline{\theta}})$.

Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and a politician of type $\theta = \overline{\theta}$ allocates $a_2 = 1$ and a politician of type $\theta = \underline{\theta}$ allocates $a_2 = 0$. All other strategies and beliefs are identical to the previous case.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \underline{\theta}$ will not deviate if:

$$\underline{\theta}q + p\underline{\theta}\tau(\frac{\overline{\theta}\pi}{\overline{\theta}\pi + (1-\pi)\underline{\theta}}, \mathbf{a}) + p(1-\underline{\theta})\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi + (1-\underline{\theta})(1-\pi)}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}) \geq$$
$$1 + p\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi + (1-\underline{\theta})(1-\pi)}, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})$$

The threshold is derived in the third case of Proposition 4.1. This profile of strategies and beliefs cannot be sustained as an equilibrium for a lower value of q.

Finally, consider candidate separating equilibria. First, note that because $\overline{\theta} > \underline{\theta}$, there cannot exist an equilibrium in which a politician of type $\theta = \underline{\theta}$ allocates $a_t = 1$ while a politician of type $\theta = \overline{\theta}$ allocates $a_t = 0$. Thus, consider equilibria in which in which a politician of type $\theta = \overline{\theta}$ allocates $a_1 = 1$ and a politician of type $\theta = \underline{\theta}$ allocates $a_1 = 0$ in the parameter spaces $q < \frac{1}{\theta}$ and $q \geq \underline{\theta}$.

- Suppose $q < \frac{1}{\theta}$: Consider the following strategy and belief profile: $\theta = \overline{\theta}$ allocates $a_1 = 1$ and a politician of type $\theta = \underline{\theta}$ allocates $a_1 = 0$; either type of politician allocates $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:
  - If the voter does not observe g_1, $\mu = \pi$;
  - Upon observation that $g_1 = q$, $\mu(g_1 = q) = 1$;

– Upon observation that $g_1 = 0$, $\mu(g_1 = 0) = \frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi+1-\pi}$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \overline{\theta}$ will not deviate if:

$$\begin{aligned}&\overline{\theta}q + p\overline{\theta}\tau(1,\mathbf{a}) + p(1-\underline{\theta})\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi+1-\pi},\mathbf{a}) + (1-p)\tau(\pi,\mathbf{a}) \geq \\ &1 + p\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi+1-\pi},\mathbf{a}) + (1-p)\tau(\pi,\mathbf{a})\end{aligned}$$

In any equilibrium in which $a_2 = 0\forall\theta$, $\tau(\mu,\boldsymbol{a})$ is equivalent for any μ, per Equation 4.9. Combined with $\overline{\theta}q < 1$ in this parameter space, this inequality never holds. Thus, this strategy and belief profile is not an equilibrium.

- Suppose $q \geq \frac{1}{\underline{\theta}}$: Consider the following strategy and belief profile: $\theta = \overline{\theta}$ allocates $a_1 = 1$ and a politician of type $\theta = \underline{\theta}$ allocates $a_1 = 1$; either type of politician allocates $a_2 = 0$. All other strategies and beliefs are identical to the previous case. The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \underline{\theta}$ will not deviate if:

$$\begin{aligned}&\underline{\theta}q + \underline{\theta}q\left[p\overline{\theta}\tau(1,\mathbf{a}) + p(1-\underline{\theta})\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi+1-\pi},\mathbf{a}) + (1-p)\tau(\pi,\mathbf{a})\right] \geq \\ &1 + \underline{\theta}q\left[p\tau(\frac{(1-\overline{\theta})\pi}{(1-\overline{\theta})\pi+1-\pi},\mathbf{a}) + (1-p)\tau(\pi,\mathbf{a})\right]\end{aligned}$$

By Lemma C.1 and $q\underline{\theta} \geq 1$ when $q \geq \frac{1}{\underline{\theta}}$, this inequality is never satisfied. Thus, this strategy and belief profile is not an equilibrium. ■

Proof of Proposition 4.2

This follows directly from Proposition 4.1.

Proof of Proposition 4.3

This proof proceeds in two sections. I first prove the existence of the equilibria characterized in Proposition 4.3, then I prove uniqueness.

Existence

First, suppose that $q < \frac{1}{\overline{\theta}}$ and consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$, and voter beliefs are as follows:

- If the voter does not observe a_1, $\mu = \pi$;
- Upon observation that $a_1 = 0$, $\mu(a_1 = 0) = \pi$;
- Off the equilibrium path, an observation that $a_1 = 1$ implies that $\mu(a_1 = 1) = 1$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. Given the voter's posterior belief $\mu = \pi$, $\tau(\pi, \mathbf{a}) = \frac{1}{2}$. $q < \frac{1}{\overline{\theta}}$ implies that $E[g_1] < a_1 \forall \theta$, such that a politician of either type allocates $a_1 = 0$. The competent type cannot profitably deviate by allocating $a_1 = 1$ because:

$$1 + p\tau(\pi, \mathbf{a}) + (1 - p)\tau(\pi, \mathbf{a}) > \overline{\theta}q + p\tau(1, \mathbf{a}) + (1 - p)\tau(\pi, \mathbf{a})$$

In this interval, $\overline{\theta}q < 1$ and $\tau(\mu, \mathbf{a}) = \frac{1}{2} \forall \mu$ when $a_1 = a_2 = 0 \forall \theta$. Since $\overline{\theta} > \underline{\theta}$, it thus holds that the incompetent type cannot profitably deviate by allocating $a_1 = 1$.

Second, suppose that $q \in \left[\frac{1}{\overline{\theta}}, \frac{2b}{\underline{\theta}2b + p\overline{\theta}}\right)$ and consider the following strategy and belief profile: a politician of type $\theta = \overline{\theta}$ allocates $a_1 = a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates

$a_1 = a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[U_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe a_1, $\mu = \pi$;
- Upon observation that $a_1 = 1$, $\mu(a_1 = 1) = 1$;
- Upon observation that $a_1 = 0$, $\mu(a_1 = 0) = 0$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**, a politician of type $\theta = \overline{\theta}$ will not deviate from $a_1 = 1$ to $a_1 = 0$ since:

$$\overline{\theta}q + (p\tau(1, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}))\,\overline{\theta}q \geq 1 + (p\tau(0, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}))\,\overline{\theta}q.$$

This inequality clearly holds for any $q \in \left[\frac{1}{\overline{\theta}}, \frac{2b}{\underline{\theta}2b + p\overline{\theta}}\right)$ since $\overline{\theta}q \geq 1$ and, by Lemma C.1, $\tau(1, \mathbf{a}) \geq \tau(0, \mathbf{a})$. A politician of type $\theta = \underline{\theta}$ cannot profitably deviate to allocate $a_1 = 1$ to increase her chances of re-election when:

$$1 + p\tau(0, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}) > \underline{\theta}q + p\tau(1, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})$$
$$\Leftrightarrow q < \frac{2b}{\underline{\theta}2b + p\overline{\theta}}$$

Third, suppose that $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b}{\underline{\theta}2b + p\overline{\theta}}\}, \frac{1}{\underline{\theta}}\right)$ and consider the following and strategy and belief profile: a politician of type $\theta = \overline{\theta}$ allocates $a_1 = a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates $a_1 = 1$ and $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe a_1, $\mu = \pi$;

- Upon observation that $a_1 = q$, $\mu(a_1 = 1) = \pi$;
- Off the equilibrium path, an observation that $a_1 = 0$ implies that $\mu(a_1 = 0) = 0$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting the equilibrium allocation strategy, **a**, a politician of type $\theta = \overline{\theta}$ will not deviate from $a_1 = 1$ to $a_1 = 0$ since:

$$\overline{\theta}q + (p\tau(\pi, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}))\overline{\theta}q > 1 + (p\tau(0, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}))\overline{\theta}q$$

This inequality holds for any $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b}{\underline{\theta}2b+p\overline{\theta}}\}, \frac{1}{\underline{\theta}}\right)$ given that $\overline{\theta}q > 1$ and, by Lemma C.1, $\tau(\pi, \mathbf{a}) \geq \tau(0, \mathbf{a})$. A politician of type $\theta = \underline{\theta}$ cannot profitably deviate by allocating $a_1 = 0$ if:

$$\underline{\theta}q + p\tau(\pi, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})) > 1 + p\tau(0, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a})$$
$$q > \frac{2b}{\underline{\theta}2b + \overline{\theta}p}$$

This inequality holds for any $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b}{\underline{\theta}2b+p\overline{\theta}}, \frac{1}{\underline{\theta}}\}\right)$ since $\pi \in (0, 1)$.

Finally, suppose that $q \geq \frac{1}{\underline{\theta}}$ and consider the following strategy and belief profile: politicians of both types allocate $a_1 = a_2 = 1$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] > E[u_2^V(c)]$; and the voter's beliefs are as follows:

- If the voter does not observe a_1, $\mu = \pi$;
- Upon observation that $a_1 = 1$, $\mu(a_1 = 1) = \pi$;
- Off the equilibrium path, an observation that $a_1 = 0$, implies that $\mu(a_1 = 0) = 0$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**, a politician of type $\theta = \underline{\theta}$ will not deviate from $a_1 = 1$ to $a_1 = 0$ since:

$$\underline{\theta}q + \tau(\pi, \mathbf{a}) > 1 + (p\tau(0, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}))\underline{\theta}q$$

This inequality holds for any $q \geq \frac{1}{\underline{\theta}}$ because $\underline{\theta}q > 1$ and $\tau(\pi, \mathbf{a}) \geq \tau(0, \mathbf{a})$ per Lemma C.1. This is sufficient to ensure that a politician of type $\theta = \overline{\theta}$ similarly does not deviate.

Uniqueness

I consider all possible pooling equilibria and then examine other possible separating equilibria. Off path beliefs are only invoked in a pooling equilibrium in which both types allocate the same a_1. In an pooling equilibrium in which both types allocate $a_1 = 0$, per the intuitive criteria, I impose the off-path belief that $\mu = 1$ upon observation of $a_1 = 1$ in any such equilibrium. There are three possible equilibria in which both types allocate $a_1 = 0$ that emerge in different regions of the parameter space, depending on a politician's second-period allocation behavior. The first is an equilibrum (the first case of Proposition 4.3), the others are not, as shown below:

- First, consider $q \in \left[\frac{1}{\overline{\theta}}, \frac{1}{\underline{\theta}}\right)$:

  Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and a politician of type $\theta = \overline{\theta} = 1$ allocates $a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

  - If the voter does not observe a_1, $\mu = \pi$;
  - Upon observation that $a_1 = 0$, $\mu(a_1 = 0) = \pi$;
  - Off the equilibrium path, an observation that $a_1 = 1$ implies that $\mu(a_1 = 1) = 1$.

The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \overline{\theta}$ will not deviate if:

$$1 + \tau(\pi, \mathbf{a}) > \overline{\theta}q + p\overline{\theta}\tau(1, \boldsymbol{a}) + p(1 - \overline{\theta})\tau(\pi, \boldsymbol{a}) + (1 - p)\tau(\pi, \boldsymbol{a})$$

However, this inequality is never satisfied since $\overline{\theta}q \geq 1$ and $\tau(1, \mathbf{a}) > \tau(\pi, \mathbf{a})$. Thus, this strategy and belief profile is not an equilibrium.

- Second, consider $q \geq \frac{1}{\underline{\theta}}$.

  Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 0$ and a politician of both types allocate $a_2 = 1$. All other beliefs and strategies are equivalent to the previous case.

  The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \overline{\theta}$ will not deviate if:

$$1 + \tau(\pi, \mathbf{a}) > \overline{\theta}q + p\overline{\theta}\tau(1, \boldsymbol{a}) + p(1 - \overline{\theta})\tau(\pi, \boldsymbol{a}) + (1 - p)\tau(\pi, \boldsymbol{a})$$

  However, this inequality is never satisfied since $\overline{\theta}q \geq 1$ and $\tau(1, \mathbf{a}) > \tau(\pi, \mathbf{a})$. Thus, this strategy and belief profile is not an equilibrium.

In an pooling equilibrium in which both types allocate $a_1 = 1$, per the intuitive criteria, I impose the off-path belief that $\mu = 0$ upon observation of $a_1 = 0$ in any such equilibrium. There are possible equilibria in which both types allocate $a_1 = 0$ that emerge in different regions of the parameter space, depending on a politician's second-period allocation behavior. One is an equilibrum (the final case of Proposition 4.2), the others are not, as shown below:

- First, suppose $q < \frac{1}{\overline{\theta}}$.

  Consider the following strategy and belief profile: politicians of both types allocate $a_1 = 1$ and a politician of either type allocates $a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

  - If the voter does not observe a_1, $\mu = \pi$;
  - Upon observation that $a_1 = 0$, $\mu(0) = \pi$;
  - Off the equilibrium path, an observation that $a_1 = 0$ implies that $\mu = 0$.

  The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, $\mathbf{a}$. A politician of type $\theta = \underline{\theta}$ will not deviate if:

  $$\underline{\theta}q + \tau(\pi, \mathbf{a}) > 1 + p\tau(0, \boldsymbol{a}) + (1-p)\tau(\pi, \boldsymbol{a})$$

  In this equilibrium, $E[g_2|\theta = \overline{\theta}] = E[g_2|\theta = \underline{\theta}] = 0$ since $a_2 = 0$ for both types. This implies that $\tau(\mu, \boldsymbol{\alpha})$ for any μ. Thus $\tau(\pi, \mathbf{a}) = \tau(0, \mathbf{a})$. Because $\underline{\theta}q < 1$ in this parameter space, the inequality is never satisfied. Thus, this strategy and belief profile is not an equilibrium.

- Second, suppose $q \in \left[\frac{1}{\overline{\theta}}, \frac{1}{\underline{\theta}}\right)$:

This equilibrium is equivalent to the third case of Proposition 4.3. This equilibrium can be sustained for any $q \in [q \geq \frac{2b}{2\underline{\theta}b+\overline{\theta}p\pi}, \frac{1}{\underline{\theta}})$.

As in the proof of Proposition 4.1, note that because $\overline{\theta} > \underline{\theta}$, there cannot exist an equilibrium in which a politician of type $\theta = \underline{\theta}$ allocates $a_t = 1$ while a politician of type $\theta = \overline{\theta}$ allocates $a_t = 0$. Thus the candidate separating equilibria are as follows.

- First, suppose $q < \frac{1}{\overline{\theta}}$.

  Consider the following strategy and belief profile: a politician of type $\theta = \overline{\theta}$ allocates $a_1 = 1$ and $a_2 = 0$ while a politician of type $\theta = \underline{\theta}$ allocates $a_1 = a_2 = 0$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:

  - If the voter does not observe a_1, $\mu = \pi$;
  - Upon observation that $a_1 = 1$, $\mu(a_1 = 1) = 1$;
  - Upon observation that $a_1 = 0$, $\mu(a_1 = 0) = 0$;

  The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \overline{\theta}$ will not deviate if:

  $$\overline{\theta}q + p\tau(1, \mathbf{a}) + (1-p)\tau(\pi, \mathbf{a}) \geq 1 + \tau(\pi, \mathbf{a}) \Leftrightarrow q \geq \frac{1}{\overline{\theta}}$$

  Thus, this profile of strategies and beliefs cannot be sustained as an equilibrium.

- Second, suppose $q \geq \frac{1}{\underline{\theta}}$ Consider the following strategy and belief profile: a politician of type $\theta = \overline{\theta}$ allocates $a_1 = a_2 = 1$ while a politician of type $\theta = \underline{\theta}$ allocates $a_1 = 0$ and $a_2 = 1$; the bureaucrat exerts effort proportional to θ in each period; the voter votes to re-elect if $E[u_2^V(i)] \geq E[u_2^V(c)]$; and the voter's beliefs are as follows:
  - If the voter does not observe a_1, $\mu = \pi$;
  - Upon observation that $a_1 = 1$, $\mu(a_1 = 1) = 1$;
  - Upon observation that $a_1 = 0$, $\mu(a_1 = 0) = 0$;

  The bureaucrat's equilibrium effort follows from inspection of Equation 4.1. The voter's choice is optimal given her posterior belief and Equation 4.9. The equilibrium second-period allocation strategies follow from inspection of Equation 4.7. Denoting equilibrium allocation strategy, **a**. A politician of type $\theta = \underline{\theta}$ will not deviate if:

  $$1 + p\tau(0, \mathbf{a}) + (1 - p)\tau(\pi, \mathbf{a}) \geq \underline{\theta}q + \tau(\pi, \mathbf{a})$$

  As $\underline{\theta}q > 1$ and $\tau(0, \mathbf{a}) < \tau(\pi, \mathbf{a})$ by Lemma C.1, this cannot be sustained as an equilibrium.

  ■

Proof of Proposition 4.4

This follows directly from the proof of Proposition 4.3.

Formal Motivation of Empirical Tests

The tests described in Table 4.1 follow directly from Propositions 4.1 and 4.3. However, in an empirical context, politicians represent a mix of first- and second-period politicians. As such, it is necessary to examine the composition of second vs. first-period incumbents in deriving predictions. Note two features of the model with respect to re-election, denoted $R(q)$:

- In the model characterized in Proposition 4.1, for any $q \geq \frac{1}{\overline{\theta}}$, $Pr(R|\theta = \overline{\theta}, q) \geq Pr(R|\theta = \underline{\theta}, q)$. For any $q < \frac{1}{\overline{\theta}}$, $Pr(R|\theta = \overline{\theta}, q) = Pr(R|\theta = \underline{\theta}) = \frac{1}{2}$.

- The rate at which incumbents are re-elected, $Pr(R|q) = \pi Pr(R|\theta = \overline{\theta}, q) + (1-\pi)Pr(R|\theta = \underline{\theta}, q)$. For any $q \notin \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(\pi-1))}\}, \frac{1}{\underline{\theta}}\right)$, the probability of re-election is $\frac{1}{2}$. For any $q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(\pi-1))}\}, \frac{1}{\underline{\theta}}\right)$, $Pr(R|q) = \frac{1}{2} - \frac{\underline{\theta}q(1-\pi)}{2b}$, $Pr(R|q) = \frac{1}{2} - \frac{\underline{\theta}q(1-\pi)}{2b}$.

Some of these re-election probabilities are cumbersome to write. For the purposes of reduced-form predictions the two bullet points suffice to illustrate the logic.

1. *Politician allocations to rents* $(1-a)$ *weakly decrease in bureaucratic quality (q):*

   This result follows from the equilibrium characterization in Proposition 4.1. In the unrestricted model, expected allocations to rents are given by:

$$E[1-a] = \begin{cases} 1 & \text{if } q < \frac{1}{\overline{\theta}} \\ \frac{1-\pi}{2} + Pr(R|\theta = \underline{\theta}, q) & \text{if } q \in \left[\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\right) \\ Pr(R|\theta = \underline{\theta}, q) & \text{if } q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right) \\ 0 & \text{if } q \geq \frac{1}{\underline{\theta}} \end{cases} \tag{C.1}$$

   Note that $Pr(R|\theta = \underline{\theta}, q)$ is greater in the second case than in the third.

   In the case in which $\overline{\theta} = 1$ and $\underline{\theta} = 0$, $E[1-a] = \frac{1-\pi}{2} + Pr(R|\theta = \underline{\theta}, q)$.

In the case in which $p = 0$, expected allocations to rents are given by:

$$E[1-a] = \begin{cases} 1 & \text{if } q < \frac{1}{\overline{\theta}} \\ 1-\pi & \text{if } q \in \left[\frac{1}{\overline{\theta}}, \frac{1}{\underline{\theta}}\right) \\ 0 & \text{if } q \geq \frac{1}{\underline{\theta}} \end{cases} \tag{C.2}$$

2. *Politicians allocate more to rents in their second term ($t = 2$) than in their first term ($t = 1$). This difference is attenuated to zero at very low and high levels of bureaucratic quality.*

   In the unrestricted model, the difference in allocations to rents, by period are given by:

$$E[1-a_2] - E[1-a_1] = \begin{cases} 0 & \text{if } q < \frac{1}{\overline{\theta}} \\ (1-\pi)(Pr(R|\theta = \underline{\theta}, q) - \frac{1}{2}) & \text{if } q \in \left[\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\right) \\ (1-\pi)(Pr(R|\theta = \underline{\theta}, q) & \text{if } q \in \left[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\right) \\ 0 & \text{if } q > \frac{1}{\overline{\theta}} \end{cases} \tag{C.3}$$

   Note that in the third case $Pr(R|\theta = \underline{\theta}, q) \leq \frac{1}{2}$.

   In the case in which $\overline{\theta} = 1$ and $\underline{\theta} = 0$:

$$E[1-a_2] - E[1-a_1] = \begin{cases} 0 & \text{if } q < \frac{1}{\overline{\theta}} \\ (1-\pi)(Pr(R|\theta = \underline{\theta}, q) - \frac{1}{2}) & \text{if } q \in \left[\frac{1}{\overline{\theta}}, \frac{1}{\underline{\theta}}\right) \\ 0 & \text{if } q > \frac{1}{\overline{\theta}} \end{cases} \tag{C.4}$$

   In the case in which $p = 0$, the diference in allocations to rents, by term, is $E[1-a_2] - E[1-a_1] = 0$.

3. *At high levels of bureaucratic quality, a voter's posterior belief (μ) is equivalent to her prior (π) upon receiving a signal that a politician allocated no funds to rents ($a = 1$).*

   This follows from Proposition 4.3 since politician actions are conveyed in the signal. For the purposes of the survey experiment, assume $p = 1$, though the qualitative results carry through for any $p > 0$.

$$\mu(a=1) = \begin{cases} 1 & \text{if } q < \frac{1}{\overline{\theta}} \text{ (Off path assumption.)} \\ 1 & \text{if } q \in \left[\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\right) \\ \pi & \text{if } q > \max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\} \end{cases} \tag{C.5}$$

In the case in which $\overline{\theta} = 1$ and $\underline{\theta} = 0$, $\mu(a = 1) = 1$.

C.2 Bureaucratic Quality Measure

C.2.1 Operationalization

The bureaucratic quality question is coded from counts of public employees in direct municipal administration according to Table C.1.

	Category (Portuguese)	Highest education	N	Value (v)
1	*Sem instrução*	Incomplete primary	N_0	$v_0 = 0$
2	*Ensino fundamental*	Complete primary	N_1	$v_1 = 1$
3	*Ensino médio*	Complete secondary	N_3	$v_2 = 2$
4	*Ensino superior*	Complete undergraduate	N_4	$v_3 = 3$
5	*Pós-graduação*	Complete post-grad	N_5	$v_4 = 4$

Table C.1: Classification of educational composition of municipal employees as reported MUNIC surveys.

The average education measure is calculated, within a survey (single year) as:

$$\text{Average education} = \frac{\sum_{c=1}^{5} N_c v_c}{\sum_{c=1}^{5} N_c} \tag{C.6}$$

Denote average education in municipality θ in year t as q_{mt}. The z-score standardization, denoted Q_{mt}, is calculated as:

$$Q_{mt} = \frac{q_{mt} - \mu_{q_{mt}}}{\sigma_{q_{mt}}} \tag{C.7}$$

where $\mu_{q_{mt}}$ denotes the mean of q_{mt} and $\sigma_{q_{mt}}$ denotes the standard deviation of $q_m t$. In estimation, all quantiles refer to the full distribution of Q_{mt} (equivalent to the quantiles of q_{mt}, *not* quantiles within the sample.

C.2.2 Description

Figure C.1 describes the bureaucratic education data graphically. Figure C.2 depicts the distribution of the raw (unstandardized) measure of bureaucratic quality over time.

Figure C.3 depicts the relationship between the set of covariates intended to adjust for variation in local labor markets. and bureaucratic quality. These provide a visualization of the fixed effects used in (non-interactive) specifications. I plot the explanatory power of these covariates in Figure C.4, showing that these covariates account less than 20% of the variation in the bureaucratic quality measure.

C.2.3 Persistence of bureaucratic quality

Measuring the persistence of the bureaucratic quality measure is important for two reasons. First, per the model, q is an exogenous parameter assumed to be outside the short-term policy options available to an incumbent. While Figure C.2 shows gradual increases in education (quality) over

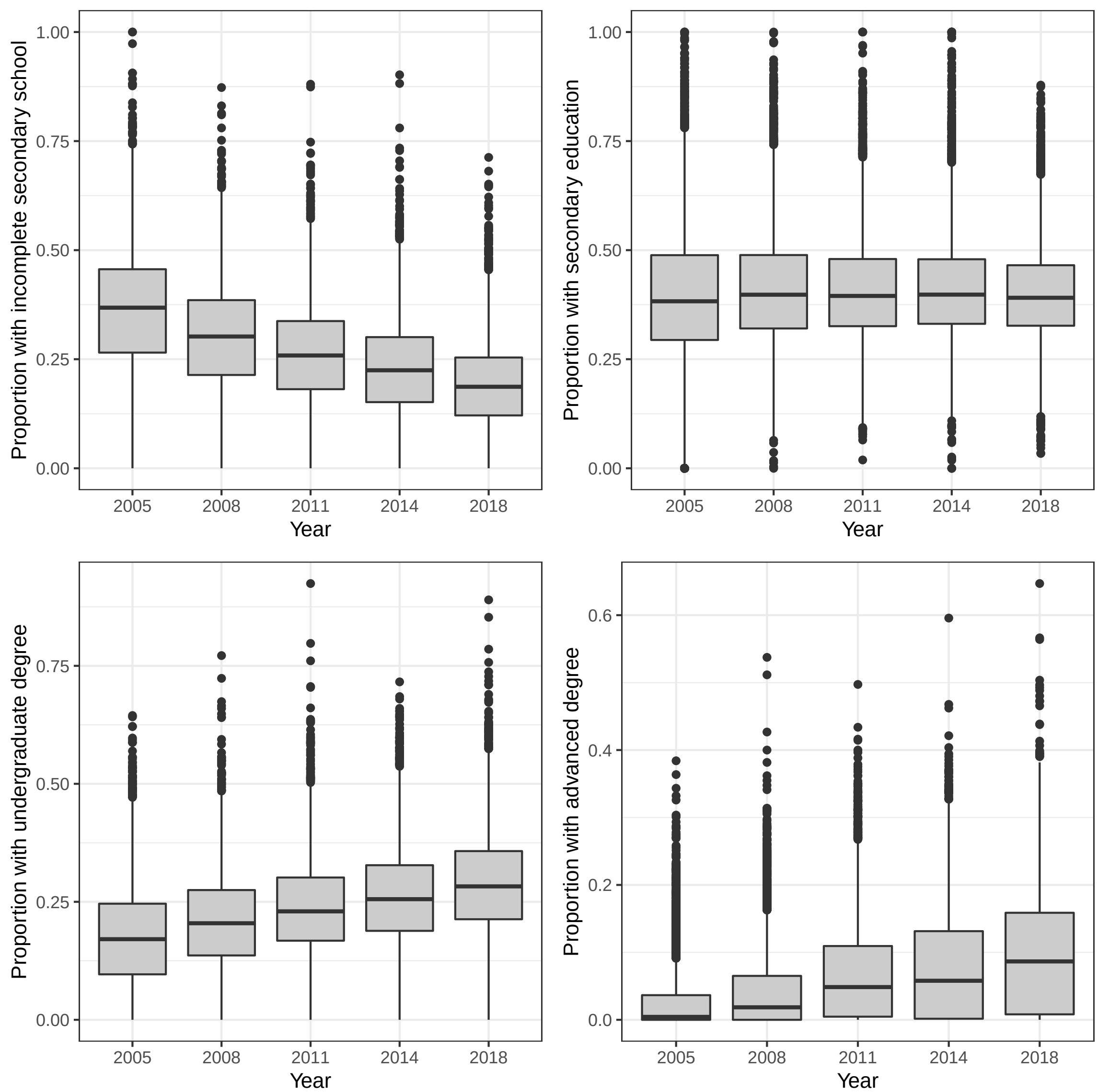


Figure C.1: Proportion of bureaucrats in categories 1-4 (per Table C.1), over time. The interquartile range (IQR) is given by the gray boxes. The confidence intervals are given by the Median $\pm\frac{1.58\text{ IQR}}{\sqrt{n}}$, where n is the number of observations.

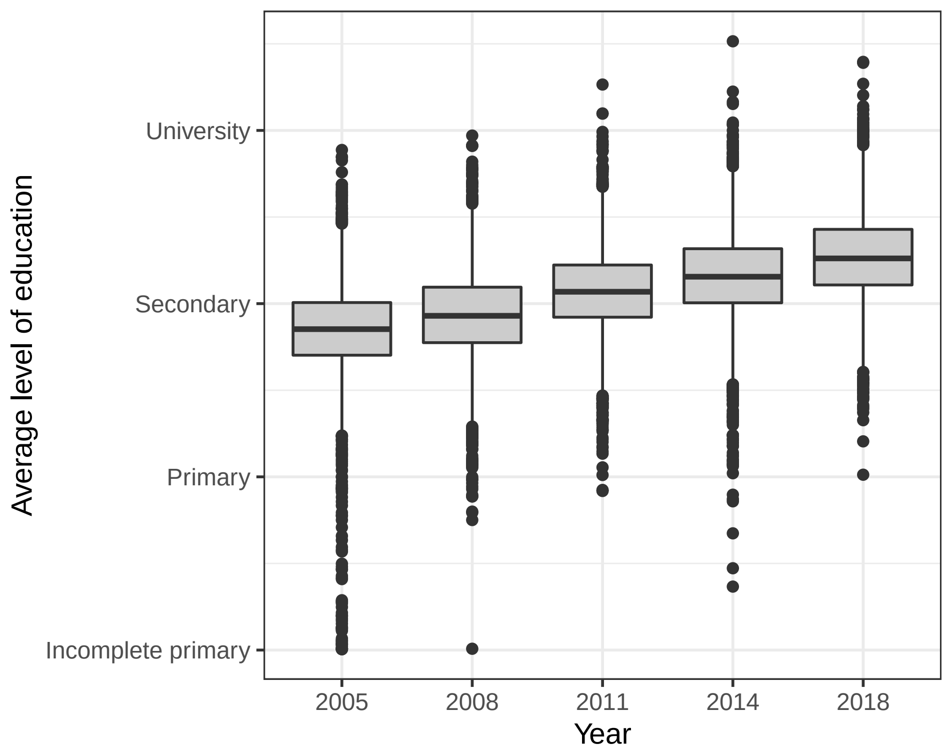


Figure C.2: Distribution of the bureaucratic quality measure (not standardized), by year. The interquartile range (IQR) is given by the gray boxes. The confidence intervals are given by the Median $\pm\frac{1.58\ \mathrm{IQR}}{\sqrt{n}}$, where n is the number of observations.

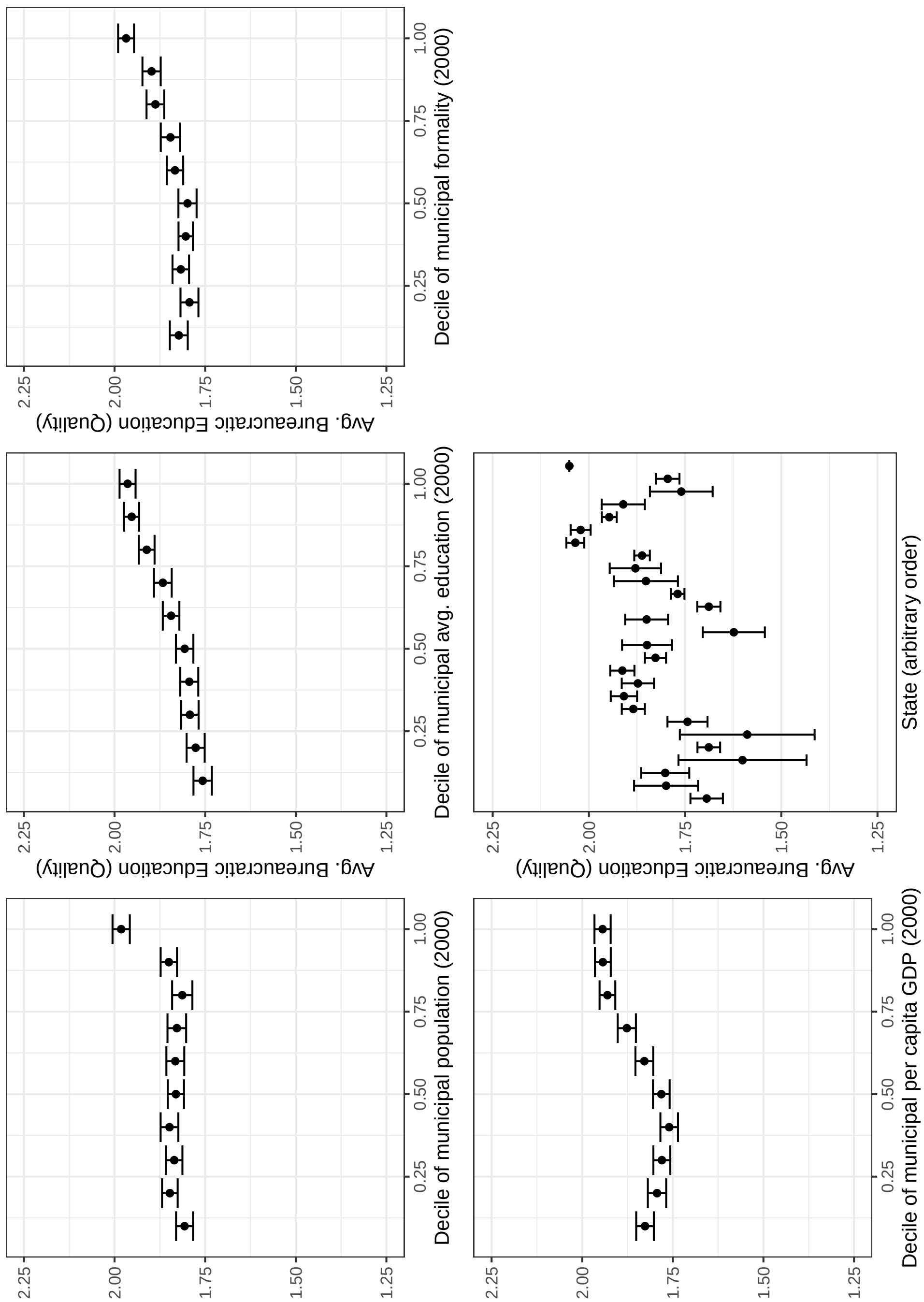


Figure C.3: Conditional means of the 2005 bureaucratic quality measure (not standardized) at deciles of municipal population, average years of education, percentage of formal employees in the workforce, and GDP per capital as well as by state. The segments represent 95% confidence intervals.

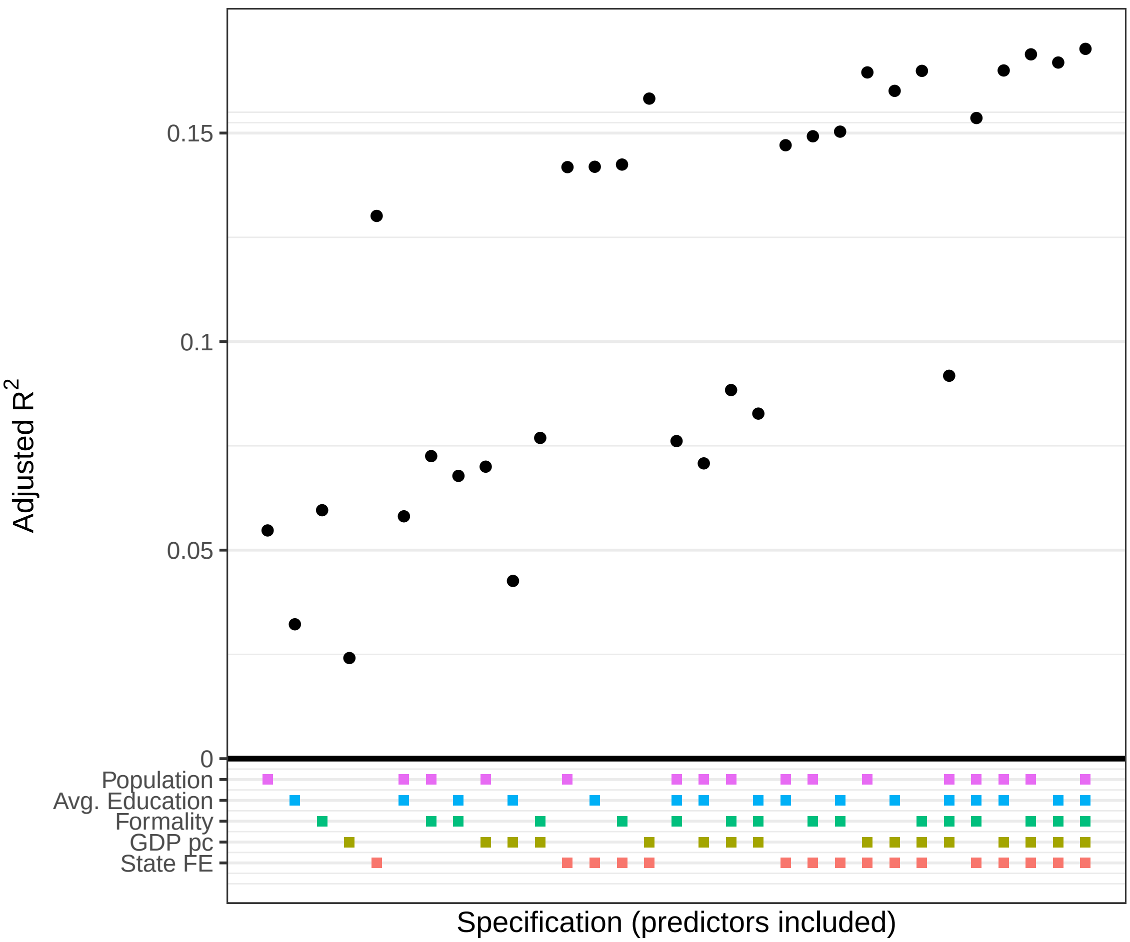


Figure C.4: This plot shows the explanatory power of the state fixed effects and binned economic covariates in predicting the bureaucratic quality measure. Each point represents the adjusted R^2 of a model regressing standardized bureaucratic quality on the set of covariates shown below the x-axis. Note that state FE have the highest predictive power and a substantial portion (>80%) of the variation in bureaucratic quality is not explained by these covariates.

Measure	Raw count/measure	Per-capita measure
Total officials in direct administration	0.977	0.851
Highest education: primary school complete	0.866	0.437
Highest education: secondary school complete	0.951	0.400
Highest education: undergraduate degree	0.975	0.473
Highest education: postgraduate degree	0.889	0.537
Average education of officials (quality)	0.574	–

Table C.2: Autocorrelation of bureaucratic education/quality measures over five waves of MUNIC. The per-capita measure of total officials uses municipal population (measured in the preceding census) as a denominator. The per-capita measures of highest education level use the number of officials in direct administration as a denominator.

time, I seek to understand whether these changes are driven by variation in the local political environment. It is important to clarify whether changes in politician (or party) yield differential changes in bureaucratic quality. Second, given the years in which education is reported in the MUNIC surveys do not align perfectly with the years in which the other data occurred/was collected, it is important to show that relative measures of bureaucratic quality are "sticky." I provide two analyses to respond to these considerations empirically.

Table C.2 reports the autocorrelation of the measures used in the construction of the bureaucratic quality measure. It indicates substantial autocorrelation across waves of the MUNIC survey for all component counts of the bureaucratic quality measure

Table C.3 conducts a first-difference analysis of changes in bureaucratic quality as a function of changes in municipal administration. Since all elections are simultaneous, the "treatments" of interest are (1) whether the mayor changes (71% of observations); and (2) whether the party of the mayor changes (68% of observations). Note that due to comparatively high rates of party switching, there are cases in which a mayor is re-elected under a different party label. I conduct a first-difference analysis of the form:

$$Q_{ms,t=1} - Q_{m,t=0} = \beta_0 + \beta_1 \text{New mayor}_m + \beta_2 \text{Different party}_m + \boldsymbol{\gamma}_s + \kappa Q_{m,t=0} + \epsilon ms$$

Table C.3 estimates this equation with OLS for each election (specifications 1-9) and then on the pooled sample. Columns 10-12 estimate this expression on the pooled sample, clustering standard errors at the municipality level. All coeficients are very small in magnitude and are generally indistinguishable from 0. In the pooled sample with covariate adjustment (Column 12), we can reject any effects outside of the [-0.003, 0.018] interval for a new mayor and outside the [-0.010, 0.011] interval for a mayor of a different party. In sum, this analysis provides no evidence that, on average, changes in leadership lead to substantive changes in bureaucratic quality.

To be sure that the effect of changing a mayor or mayoral party is not obscured by examining only mean shifts, I plot the ECDFs of the differenced bureaucratic quality outcome by each political "treatment" in Figure C.5. There is no evidence of effects on the variance.

	Δ Bureaucratic Quality					
	2008-2011			2011-2014		
	(1)	(2)	(3)	(4)	(5)	(6)
Change in mayor	-0.011	-0.007	-0.004	0.008	0.009	0.009
	(0.009)	(0.009)	(0.007)	(0.012)	(0.012)	(0.010)
Change in party	0.013	0.003	0.002	-0.017	-0.019^{*}	-0.016^{*}
	(0.009)	(0.009)	(0.008)	(0.011)	(0.011)	(0.009)
State FE		✓	✓		✓	✓
Lagged bureaucratic quality			✓			✓
DV Mean, no change	0.137	0.137	0.137	0.084	0.084	0.084
DV St. Dev, no change	0.261	0.261	0.261	0.251	0.251	0.251
Adj. R^2	0.000	0.026	0.360	0.000	0.003	0.255
Num. obs.	4932	4932	4932	4719	4719	4719
Election year	2008	2008	2008	2012	2012	2012
	2014-2018			Pooled		
	(7)	(8)	(9)	(10)	(11)	(12)
Change in mayor	-0.014	-0.015	-0.003	-0.014^{**}	-0.011^{*}	0.008
	(0.012)	(0.012)	(0.010)	(0.006)	(0.006)	(0.005)
Change in party	0.003	0.002	-0.001	0.002	-0.002	0.001
	(0.011)	(0.011)	(0.010)	(0.006)	(0.006)	(0.005)
State FE		✓	✓		✓	✓
Lagged bureaucratic quality			✓			✓
DV Mean, no change	0.104	0.104	0.104	0.109	0.109	0.109
DV St. Dev, no change	0.251	0.251	0.251	0.255	0.255	0.255
Adj. R^2	-0.000	0.003	0.319	0.000	0.007	0.293
Num. obs.	4362	4362	4362	14013	14013	14013
N Clusters				5293	5293	5293
Election year	2016	2016	2016	All	All	All

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.3: First difference analysis of the effects of changing a mayor or partisan affiliation of the mayor in an election on bureaucratic quality. The cross-sectional specifications use heteroskedasticity-robust standard errors and the panel specification clusters standard errors at the municipal level.

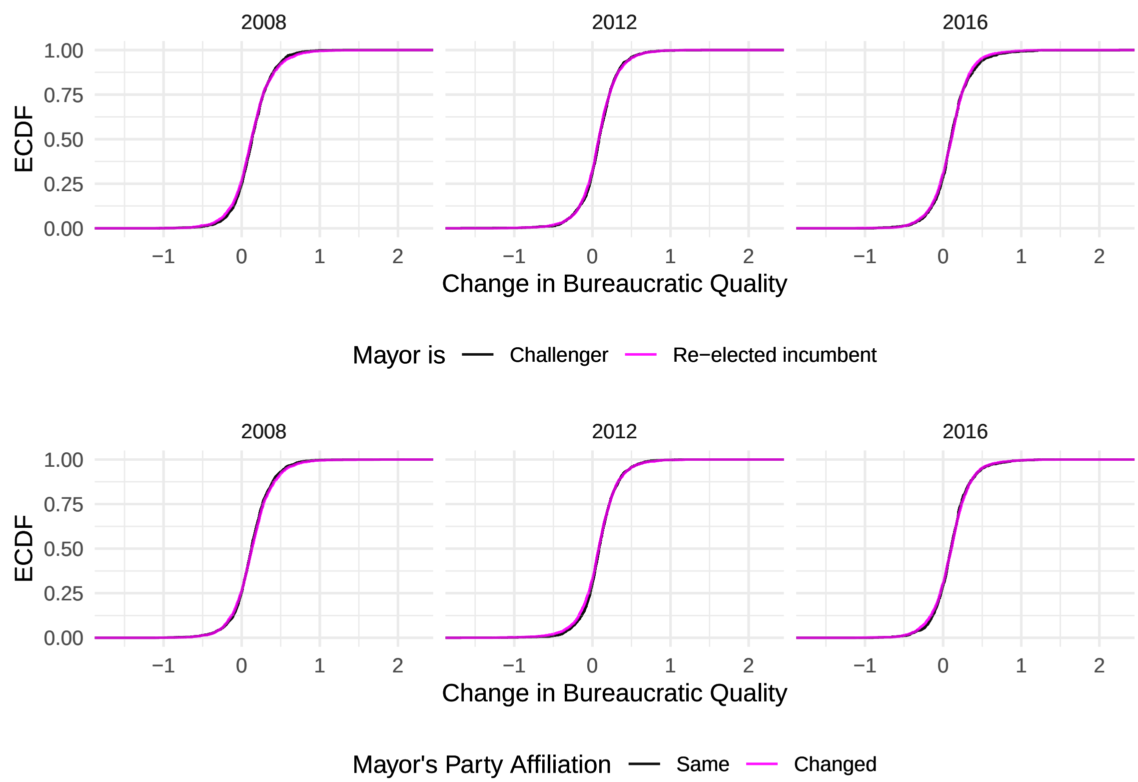


Figure C.5: These graphs plot the ECDF of differences in bureaucratic quality for each of the political "treatments" examined in Table C.3.

Finally, I examine correlation between bureaucratic quality and the presence of community radio. Community radio is the the medium through which information from audit investigations is purported to diffuse (Ferraz and Finan, 2008). Note that, in general, existing evidence suggests that community radio simply diffuses informational signals if they emerge, i.e., the results of local audits. There is not evidence that the presence of a community radio station alone increments the probability of revelation (p in the model).

I gather data on community radio from ANATEL, Brazil's National Telecommunications Agency. I use ANATEL's database of historical licensing of FM radio stations to collect the data.[1] I examine the radio stations that were licensed on December 31 of the preceding year.

To ensure that bureaucratic quality is not simply capturing community radio presence, I examine the association between bureaucratic quality and radio presence in each year that I study. Figure C.6 plots the association between bureaucratic quality and community radio presence in 2004/2005 and 2011. The top row reveals a positive correlation between bureaucratic quality and municipal radio presence. However, when examining a residualized measure of the radio presence that partials out state indicators and the demographic/economic indicators used in all models, there is no association between bureaucratic quality and community radio. Table C.4 provides a more formal test of the relationship in the scatter plots, reaching a similar conclusion.

[1] See http://sistemas.anatel.gov.br/se/public/view/b/srd.php for data.

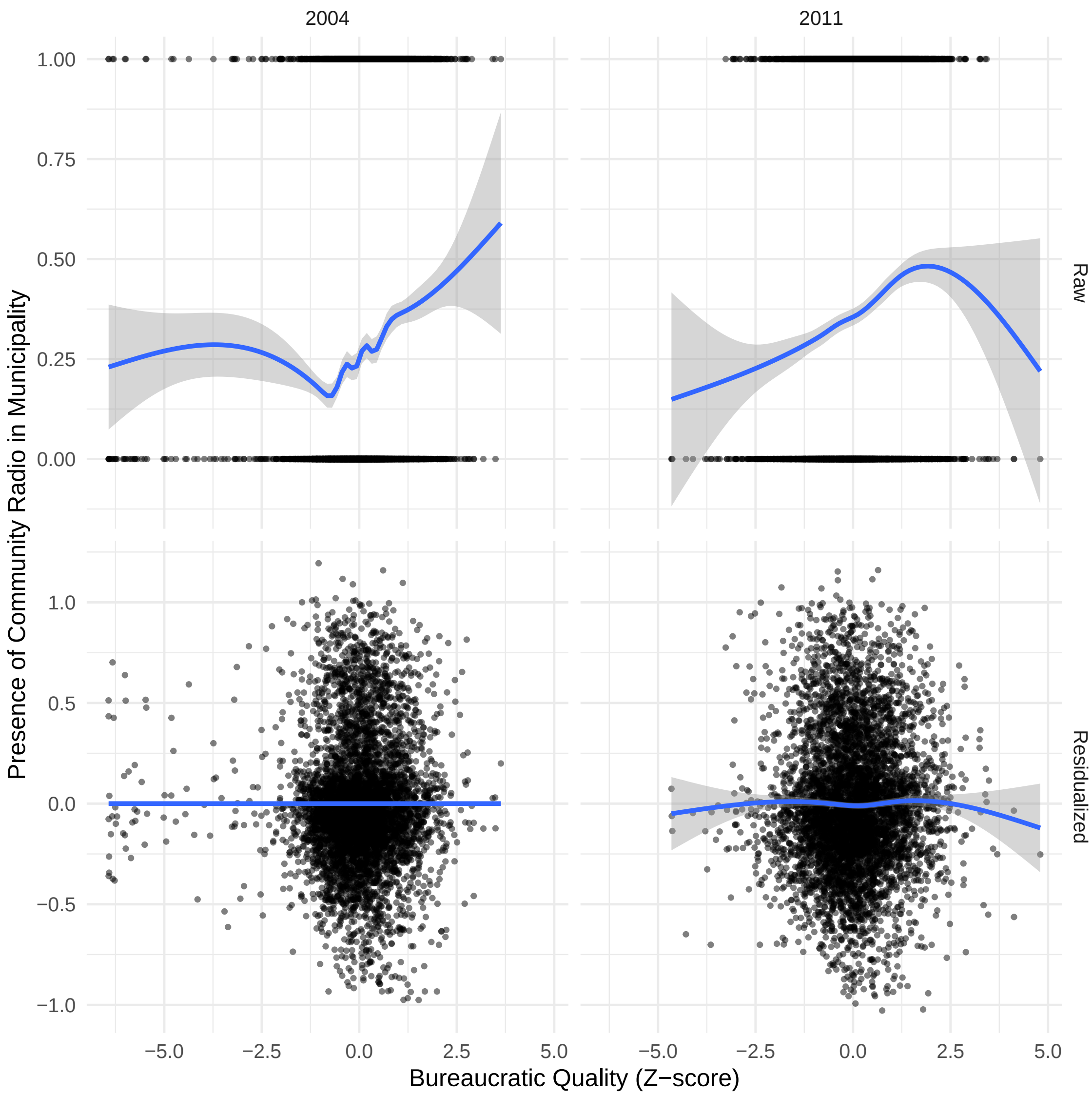


Figure C.6: This graph plots the correlation between bureaucratic quality and presence of community radio in 2004/2005 (left) and 2011 (right), the years used in the empirical tests. The bottom panel looks at a residualized presence of community radio, with the set of economic covariates (municipal population, education, formality, and GDP per capita decile bins) and state fixed effects. While there is a positive association between the raw measures of bureaucratic quality, this association is absent with the standard set of covariates used in this paper.

	Community radio in municipality (Indicator)			
	(1)	(2)	(3)	(4)
Bureaucratic Quality (z-score)	0.053***	−0.000	0.059***	−0.000
	(0.007)	(0.005)	(0.007)	(0.005)
Sample (year)	2004	2004	2011	2011
State FE		✓		✓
Demographic covariates		✓		✓
Adj. R^2	0.015	0.451	0.015	0.439
Num. obs.	5349	5347	5230	5230
RMSE	0.437	0.326	0.477	0.360

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.4: Association between bureaucratic quality and community radio presence. The demographic covariates include municipal population, education, formality, and GDP per capita decile bins. Heteroskedasticity-robust standard errors in parentheses.

C.3 Bureaucratic Quality and Allocation to Rents

C.3.1 Plots of Raw Data

The bivariate relationship between bureaucratic quality (Z-score) and share of funds spent in a corrupt manner are graphed in Figure C.7.

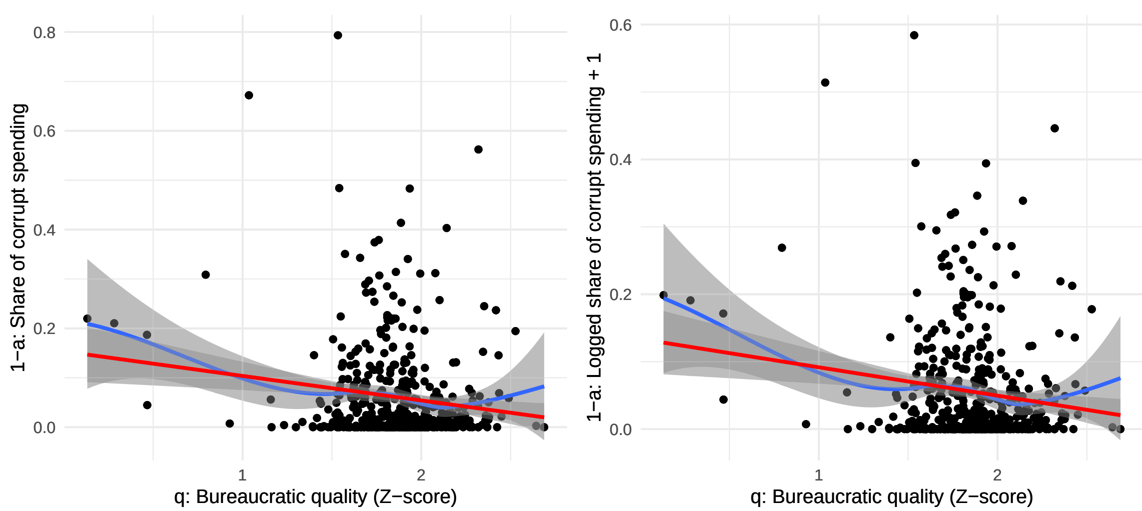


Figure C.7: Scatter plot depicting bureaucratic quality and the share of audited funds spent in a corrupt manner. These graphs plot the raw data from Table 4.2.

	Share of spending								
	Granted to corrupt bids			Misallocated			Spent on overbudget projects		
Bureaucratic quality (Z-score)	-0.008	-0.007	-0.011^{*}	-0.007^{**}	-0.007^{**}	-0.008^{*}	0.001	0.001	0.001
	(0.005)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)
DV Mean	0.040	0.040	0.040	0.021	0.021	0.021	0.001	0.001	0.001
DV Std. Dev.	0.079	0.079	0.079	0.054	0.054	0.054	0.01	0.01	0.01
Range, DV	[0,0.672]	[0,0.672]	[0,0.672]	[0,0.584]	[0,0.584]	[0,0.584]	[0,0.143]	[0,0.143]	[0,0.143]
Adj. R^2	0.010	0.061	0.074	0.013	0.056	0.059	0.002	0.001	0.009
Num. obs.	448	448	448	448	448	448	448	448	448

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.5: Decomposition of sources of corrupt spending in Table 4.2. Heteroskedasticity-robust standard errors in parentheses.

C.3.2 Decomposition of corrupt spending

One potential concern with the results in Table 4.2 is that low bureaucratic quality corresponds to worse record-keeping that would manifest in audits as corrupt spending. If this were the case, we may expect similar effects across types of malfeasant spending. This is not the case when we decompose the sources of rents in Table C.5. Increases in bureaucratic quality correlate most strongly with reductions in misallocated spending.

C.3.3 No heterogeneity by community radio presence

Diffusion of pre-2004 audit reports was believed to be facilitated by the presence of a municipal radio station in Brazilian municipalities. Note that Ferraz and Finan (2008) show that community radio magnified the electoral effects of revelation of audit information. They do not find that radio stations alone make voters more likely to sanction politicians. This section evaluates whether the presence of a local radio station influences the a politician's allocation behavior when audits were not yet anticipated. If radios do not alone increase p (absent audits) as in Ferraz and Finan (2008), then there should be no difference in allocation behavior as a function of the presence of a community radio station.

I collect historic FM radio station registrations from ANATEL and create an indicator measuring whether each municipality had an FM radio station registered in 2003. Table C.6 finds no

	Share of corrupt spending			Log(Share of corrupt spending + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Bureaucratic Quality (Z-score)	-0.015^{**}	-0.015^{**}	-0.017^{**}	-0.013^{**}	-0.013^{**}	-0.014^{**}
	(0.006)	(0.006)	(0.008)	(0.005)	(0.006)	(0.007)
Radio	-0.005	0.007	0.009	-0.005	0.006	0.008
	(0.011)	(0.012)	(0.016)	(0.009)	(0.010)	(0.013)
Bureaucratic Quality×Radio	0.002	0.001	-0.002	0.002	0.001	-0.002
	(0.013)	(0.012)	(0.012)	(0.011)	(0.010)	(0.010)
State FE		✓	✓		✓	✓
Lottery FE		✓	✓		✓	✓
Demographic controls			✓			✓
Outcome Range	[0,0.794]	[0,0.794]	[0,0.794]	[0,0.584]	[0,0.584]	[0,0.584]
Outcome Mean	0.062	0.062	0.062	0.056	0.056	0.056
Outcome Std. Dev.	0.10	0.10	0.10	0.085	0.085	0.085
Adj. R^2	0.014	0.081	0.097	0.015	0.092	0.108
Num. obs.	448	448	448	448	448	448

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.6: Heterogeneity in the association of bureaucratic quality and corruption as a function of community radio presence. 177 of the sampled communities registered community radios in December 2003 4.2. Heteroskedasticity-robust standard errors in parentheses.

heterogeneity by radio presence. I interpret this as evidence that incumbents did not differentially anticipate revelation of performance information as a function of radio presence/absence when making allocations.

C.4 First-term vs. Second-term Allocation to Rents

Given the estimator in Equation 4.11, the quantity of interest is $\widehat{\beta}_1 + \widehat{\beta}_3 Q_m$. Table C.7 suggests that this quantity is positive at low quantiles of bureaucratic quality but indistinguishable from 0 at high quantiles. The estimates of β_1 are consistently positive and statistically significant. The significance of the interaction term varies, though its sign is consistently negative. Ultimately the inference that I draw is on the quantity $\widehat{\beta}_1 + \widehat{\beta}_3 Q_m$, not simply $\widehat{\beta}_3$.

To decompose the compound mechanism behind term effects, I use a RDD in an attempt to vary the composition of the second period mayors by varying bandwidths. As I am interested in average differences, as opposed to CATEs at the threshold where the margin of victory is equal to zero, I use zero-degree polynomials in contrast to increasingly standard practice in RDs. I estimate

	Share of corrupt spending				
	(1)	(2)	(3)	(4)	(5)
LINEAR BUREAUCRATIC QUALITY MEASURE (Z-SCORE)					
Second term	0.021**	0.021**	0.018*	0.022**	0.022**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Bureaucratic quality (Z-score)	-0.007	-0.009	-0.015*	-0.007	-0.009
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)
Second term × BQ	-0.019	-0.012	-0.007	-0.018	-0.016
	(0.012)	(0.014)	(0.014)	(0.013)	(0.015)
BUREAUCRATIC QUALITY TERCILES					
Second term	0.050**	0.043*	0.034*	0.023**	0.023**
	(0.022)	(0.022)	(0.020)	(0.010)	(0.011)
Bureaucratic quality, tercile 2	0.011	0.007	0.003	0.010	0.009
	(0.015)	(0.014)	(0.015)	(0.015)	(0.016)
Bureaucratic quality, tercile 3	-0.017	-0.021	-0.035*	-0.018	-0.021
	(0.015)	(0.016)	(0.021)	(0.015)	(0.017)
Second term × BQ tercile 2	-0.052**	-0.042	-0.033	-0.053**	-0.053*
	(0.026)	(0.026)	(0.024)	(0.026)	(0.028)
Second term × BQ tercile 23	-0.029	-0.017	-0.009	-0.028	-0.024
	(0.026)	(0.026)	(0.027)	(0.026)	(0.033)
BUREAUCRATIC QUALITY QUARTILES					
Second term	0.053**	0.045*	0.035	0.023**	0.022**
	(0.026)	(0.027)	(0.025)	(0.010)	(0.011)
Bureaucratic quality, quartile 2	0.008	0.012	0.010	0.008	0.012
	(0.017)	(0.020)	(0.020)	(0.017)	(0.019)
Bureaucratic quality, quartile 3	-0.009	-0.018	-0.030	-0.011	-0.011
	(0.017)	(0.018)	(0.021)	(0.017)	(0.018)
Bureaucratic quality, quartile 4	-0.01	-0.020	-0.035	-0.016	-0.015
	(0.017)	(0.020)	(0.025)	(0.018)	(0.021)
Second term × BQ quartile 2	-0.046	-0.043	-0.034	-0.051*	-0.054*
	(0.031)	(0.032)	(0.030)	(0.030)	(0.031)
Second term × BQ quartile 3	-0.032	-0.018	-0.005	-0.031	-0.029
	(0.031)	(0.033)	(0.031)	(0.031)	(0.033)
Second term × BQ quartile 4	-0.044	-0.033	-0.021	-0.043	-0.043
	(0.030)	(0.032)	(0.033)	(0.031)	(0.038)
State FE		✓	✓		
Lottery FE		✓	✓	✓	✓
Demographic covariates			✓		✓
Covariate × term interactions				✓	✓
Num. obs.	448	448	448	448	448

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.7: Conditional associations between politician term and rent allocation, by levels of bureaucratic quality. The interactive specifications in Columns 4 and 5 use the estimator proposed in Lin (2013). All models are estimated by OLS with heteroskedasticity-robust standard errors in parentheses.

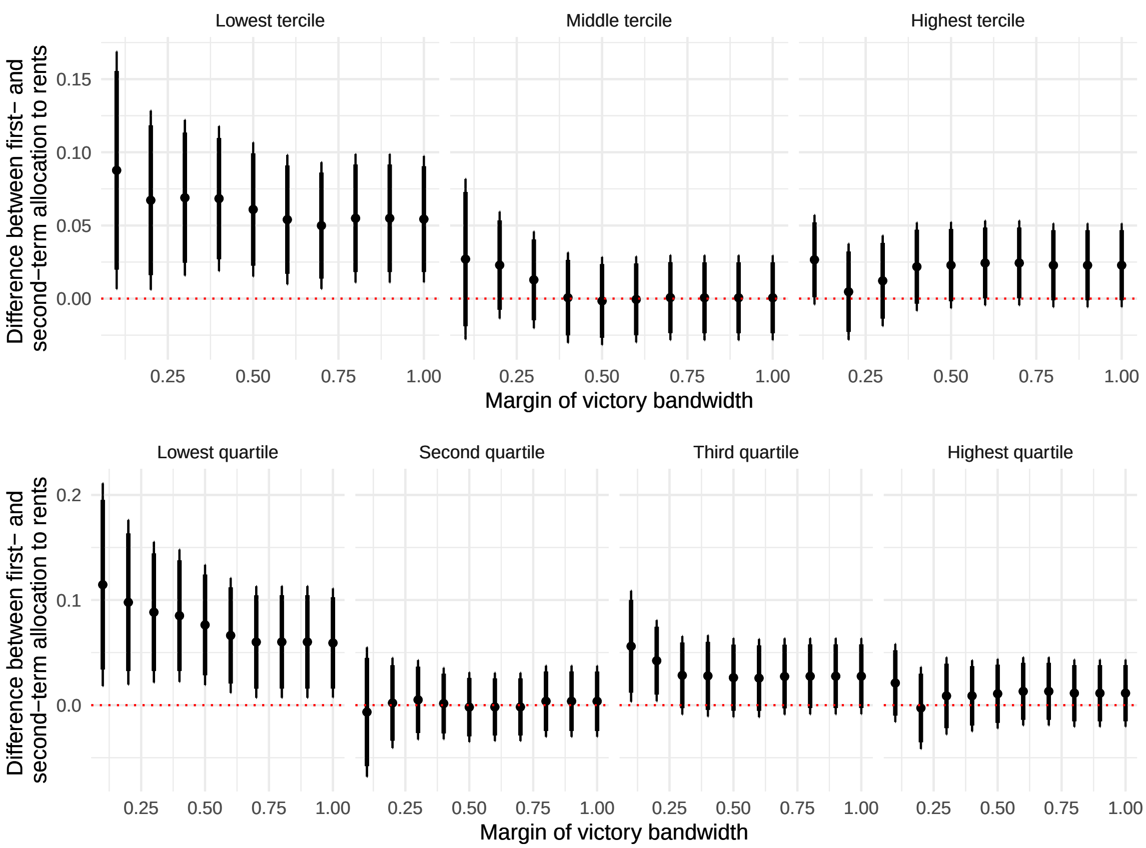


Figure C.8: Results of an experimental analogue (e.g. polynomial degree zero) of a RD specification at varying bandwidths. At low bandwidths, incompetent types are more common among re-elected politicians and differences by term are exaggerated.

Equation 4.11 at different bandwidths in terms of the 2000 margin of victory, starting with 0.1, which is smaller than the bandwidth selected in (Ferraz and Finan, 2011).[2] At smaller bandwidths, incompetent types should theoretically represent a larger share of the second-term politicians. Since these are the mayors predicted to extract rents in their second term, the marginal effect of term should be larger at small bandwidths, but only at low levels of bureaucratic quality. This is consistent with the point estimates (and differences between the narrowest and widest bandwidths) in Figure C.8.

[2]To maintain a common set of covariates across bandwidths, I omit the covariates except for lottery fixed effects in this analysis. The estimates are substantively similar with covariates but I lack degrees of freedom to estimate effects at the narrowest bandwidths.

C.5 Survey Experimental Test of Voter Updating

This paper uses a subset of treatment conditions from Weitz-Shapiro and Winters (2016*a*) and Winters and Weitz-Shapiro (2016). The full seven-arm design is enumerated in Table C.8. Because "clean" and "corrupt" are both experimental manipulations of interest, I omit treatment conditions that are not fully crossed for both types of information. I use the control (no information) condition as a measure of priors.

Arm							
Corruption Information	None	Clean	Corrupt				
Source of Information	None	Unspecified	Unspecifed	Opposition Party		Federal Audit	
Implicated Official	–	Mayor	Mayor	Mayor	Municipal official	Mayor	Municipal official
Analyzed in extension	✓	✓	✓				
N per arm:	286	286	286				

Table C.8: Design and specification of treatment conditions utilized in extension of the survey experiment.

The vignette used as the material for the three treatments of interest is quoted in Table C.9.

Arm	Vignette Text
Control	"Imagine that you live in a neighborhood similar to your own but in a different city in Brazil. Let's call the mayor of that hypothetical city in which you live Carlos. Imagine that Mayor Carlos is running for reelection. During the four years that he has been mayor, the municipality has experienced a number of improvements, including good economic growth and better health services and transportation." (Weitz-Shapiro and Winters, 2016*a*: p.66)
Clean	Control text + "Also, it is well known in the city that Mayor Carlos has not accepted any bribes when awarding city contracts." (Weitz-Shapiro and Winters, 2016*a*: p.66, emphasis added).
Corrupt	Control text + "Also, it is well known in the city that Mayor Carlos has accepted bribes when awarding city contracts." (Weitz-Shapiro and Winters, 2016*a*: p.66, emphasis added).

Table C.9: Vignette text for each treatment condition.

Sampling and Blocking

Per Weitz-Shapiro and Winters (2016*b*), the sampling procedure for cities and individuals was as follows:

> "140 cities were sampled using a probability-proportional-to-size (PPS) method within 25 strata that are defined by 25 of Brazil's 27 states. (The survey rotates on a monthly basis among three small states in the northern region of the country.) Census tracts were selected using PPS with stratification across zones of major metropolitan areas. Enumerators recruited individual respondents in public or semi-public places according to a quota scheme designed to produce a representative sample of the national population in terms of age, gender, and employment characteristics (sector of the economy and employment status)." (Weitz-Shapiro and Winters, 2016*b*: p. 4)

Because larger cities are more likely to be chosen when municipal sampling is proportional to population and larger cities have higher average bureaucratic quality (see Figure C.3), sampled municipalities have a slightly higher level of bureaucratic quality, as depicted in Figure C.9. Importantly, however, there is support across most of the distribution of bureaucratic quality. Table C.10 confirms that adjusting for municipal population eliminates this imbalance, consistent with the account of municipal sampling.

The survey experiment blocks assignment to the experimental manipulations on municipality and maintains equal probabilities of assignment in each municipality.

Robustness and Extensions

This section provides three extensions of the analysis reported in the paper, as follows:

- 129/140 municipalities in the survey experimental sample recorded bureaucratic education in 2011. I also constructed an predicted measure from an additional 10 municipalities that

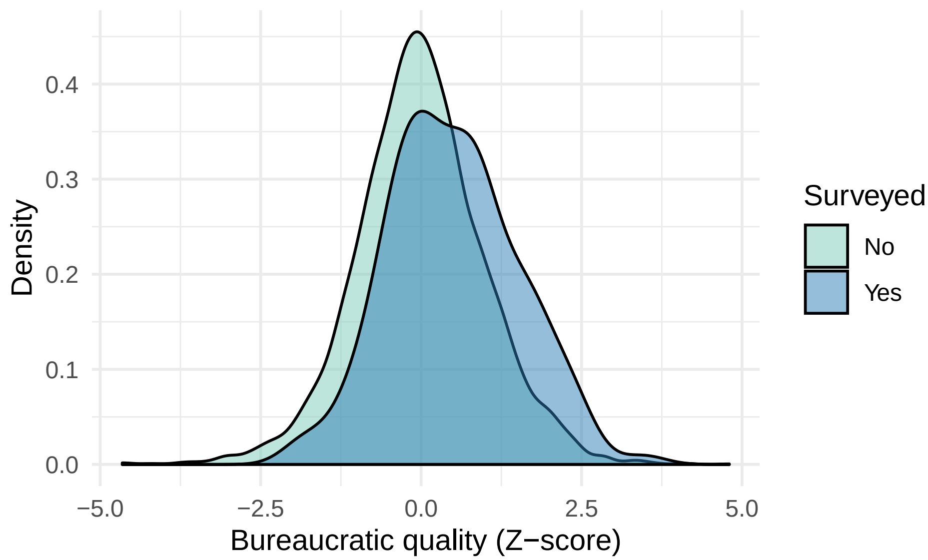


Figure C.9: Distribution of bureaucratic quality in sampled and unsampled municipalities.

	Municipality in survey sample			
	(1)	(2)	(3)	(4)
Bureaucratic quality (z-score)	0.012^{***}	0.001		
	(0.002)	(0.002)		
Bureaucratic quality w/ imputation (z-score)			0.012^{***}	0.001
			(0.002)	(0.002)
Population percentile		✓		✓
Adj. R^2	0.006	0.223	0.006	0.213
Num. obs.	5230	5230	5507	5507

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.10: Municipal sampling in survey experiment. Adjustment for municipal population accounts for differences in bureaucratic quality in sampled and non-sampled municipalities.

recorded bureaucratic education in 2008. Figure C.10 replicates Figure 4.3 from the main analysis with this slightly larger sample of municipalities.

- Figure C.11 replicates the analysis in Figure 4.3 with an alternate dependent variable, a 4-item scale measuring who voters intended to vote for in the hypothetical election.

- Figure C.12 disaggregates the result in Figure 4.3 by respondent education/political knowledge. While the subgroups reduce sample sizes and add noise, we do not see substantial differences in updating behavior across the two subgroups.

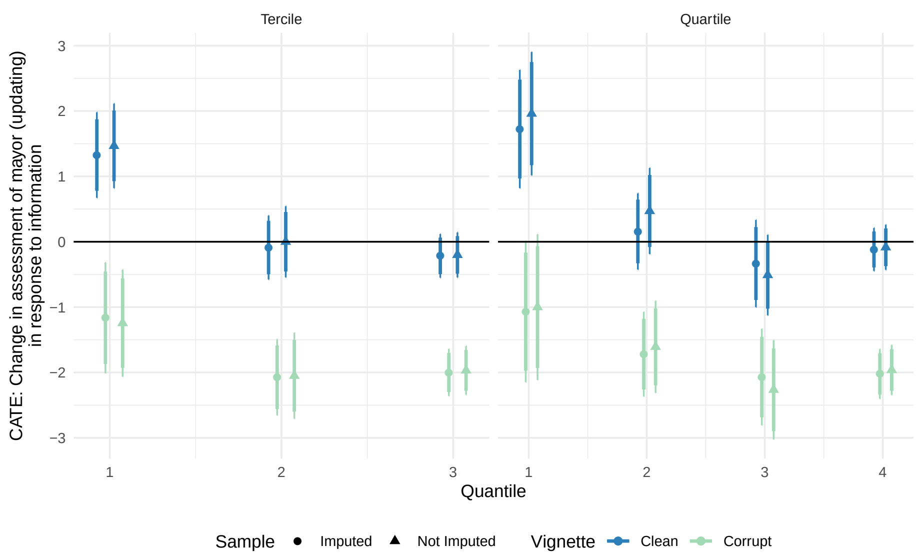


Figure C.10: Main results from Figure 4.3 replicated alongside results from a larger number of sampled municipalities with imputed measures of bureaucratic quality predicted from 2008 levels.

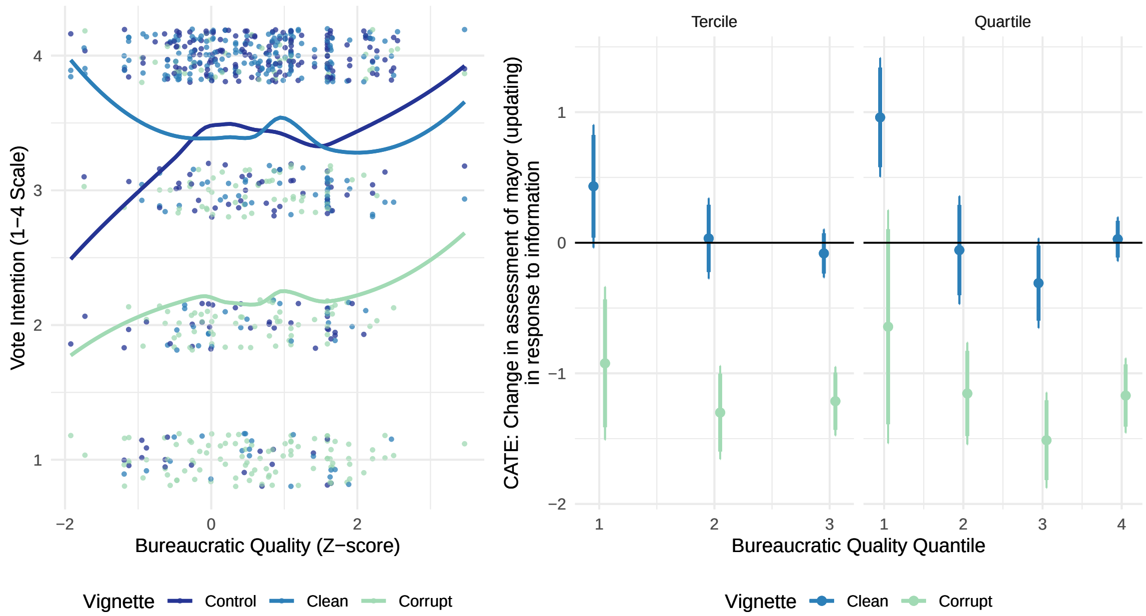


Figure C.11: Recreation of the Figure 4.3 using a "vote intent" outcome that is preferred by Weitz-Shapiro and Winters (2016*a*). All substantive results are identical, though ceiling effects appear to be of greater concern here than with the main measure.

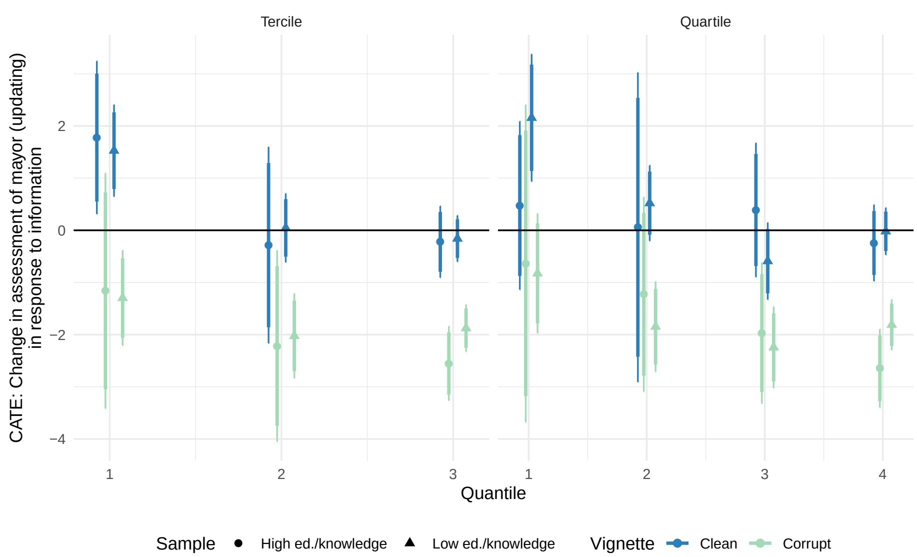


Figure C.12: Disaggregating results by subjects with high education or high political knowledge ($n = 216$) versus not ($n = 642$) reveals little heterogeneity in updating by respondent characteristics.

	Re-Elected					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of corrupt spending	-0.75^{**}	-0.80^{***}	-0.71^{*}			
	(0.29)	(0.29)	(0.33)			
Any corrupt spending				-0.14^{*}	-0.12	-0.16^{*}
				(0.08)	(0.08)	(0.09)
	Vote Share$_{2004}$ - Vote Share$_{2000}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of corrupt spending	-0.20^{*}	-0.16^{*}	-0.20^{*}			
	(0.11)	(0.09)	(0.11)			
Any corrupt spending				-0.03	-0.04^{*}	-0.06^{**}
				(0.03)	(0.02)	(0.02)
2000 Vote share		✓	✓		✓	✓
2000 Number of candidates		✓	✓		✓	✓
State FE			✓			✓
Community radio indicator			✓			✓
Lottery Round FE			✓			✓
Num. obs.	186	186	186	186	186	186

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.11: Replication of Ferraz and Finan (2008) for the audited municipalities with an incumbent running for re-election using the corruption metrics used in this paper.

C.6 Effects of Information Provision

C.6.1 Corruption and electoral performance in the sample

Table C.11 replicates Ferraz and Finan (2008) for the sample of audited municipalities with an incumbent running. Consistent with that paper, electoral performance is decreasing in corruption. Note that this table is intended only to demonstrate that the sign of the association between (endogenous) corruption and vote share is consistent with predictions.

C.6.2 Bureaucratic quality and electoral performance of incumbents

Table C.12 examines the relationship between bureaucratic quality and re-election. I find modest evidence that incumbents in places with higher levels of bureaucratic quality perform better consistent with the idea that any incumbency disadvantage occurs in lower quantiles of the sample. The difference in vote share is generally better powered than the binary indicator for re-election.

	Re-Elected					
	(1)	(2)	(3)	(4)	(5)	(6)
Bureaucratic quality (Z-score)	0.013	0.015	0.005			
	(0.010)	(0.010)	(0.011)			
Bureaucratic quality, tercile 2				0.003	0.011	−0.001
				(0.026)	(0.025)	(0.026)
Bureaucratic quality, tercile 3				0.035	0.039	0.005
				(0.025)	(0.025)	(0.028)
	Vote Share_{2004} - Vote Share_{2000}					
	(1)	(2)	(3)	(4)	(5)	(6)
Bureaucratic quality (Z-score)	0.007^{**}	0.006^{**}	0.003			
	(0.003)	(0.003)	(0.003)			
Bureaucratic quality, tercile 2				−0.000	−0.006	−0.010
				(0.008)	(0.007)	(0.008)
Bureaucratic quality, tercile 3				0.017^{**}	0.015^{**}	0.003
				(0.008)	(0.007)	(0.008)
2000 Vote share		✓	✓		✓	✓
2000 Number of candidates		✓	✓		✓	✓
State FE			✓			✓
Demographic controls (binned)			✓			✓
Num. obs.	2228	2228	2228	2228	2228	2228

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.12: Bureaucratic quality and electoral performance among municipalities with an incumbent running.

C.6.3 The effect of information

Here, we are interested in the effect of being audited (regardless of the outcome of the audit) on the probability of re-election. Denoting re-election by $R \in \{0, 1\}$ and assignment to a randomized audits as $Z = 1$, the ATE is defined as:

$$E[R|Z = 1] - E[R|Z = 0]$$

I provide a slight adaption to the model to account for the research design introduced by the audits. I assume that after voters observe (resp. does not observe g_1), they probabilistically receive another signal that reveals the politician's allocation if assigned to treatment, $Z = 1$. Specifically, with probability p_n, citizens observe the signal realization N via observation of public goods per the model characterized in Proposition 4.1. When assigned to treatment, with probability p_a, a citizen observes the realization $A \in \{0, 1\}$, which is given by an exogenous report of the politician's first period allocation $a_1 \in \{0, 1\}$. I assume that $N \perp A$. I assume that the prior may be heterogeneous, but is an accurate assessment of the proportion of types in the candidate pool.

Consider the four cases of the Equilibrium in Proposition 4.1:

- $q < \frac{1}{\theta}$: Here, the posterior belief, upon realization of N is $\mu = \pi$. Upon realization of A, the posterior belief is similarly $\mu = \pi$. It is thus trivial to show that:

$$\begin{aligned}
E[R|Z=1] &= \int_\pi \big[\pi\left(p_a p_n \tau(\pi,\mathbf{a}) + p_a(1-p_n)\tau(\pi,\mathbf{a}) + p_n(1-p_a)\tau(\pi,\mathbf{a}) + (1-p_a)(1-p_n)\tau(\pi,\mathbf{a})\right) + \\
&\quad (1-\pi)\left(p_a p_n \tau(\pi,\mathbf{a}) + p_a(1-p_n)\tau(\pi,\mathbf{a}) + p_n(1-p_a)\tau(\pi,\mathbf{a}) + (1-p_a)(1-p_n)\tau(\pi,\mathbf{a})\right)\big] d\pi \\
&= \int_\pi \tau(\pi,\mathbf{a}) d\pi \\
&= \frac{1}{2} \\
E[R|Z=0] &= \pi\left(p_n \tau(\pi,\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\right) + (1-\pi)\left(p_n\tau(\pi,\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\right) d\pi \\
&= \int_\pi \tau(\pi,\mathbf{a}) d\pi \\
&= \frac{1}{2}
\end{aligned}$$

- $q \in [\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))})$: Here, the posterior belief upon realization of N is $\mu = 1$ if public goods are observed, and $\mu = \frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi}$ if public goods are not observed. The posterior beliefs upon realization of A are $\mu = 1$ if $a = 1$ and $\mu = 0$ if $a = 0$.

$$\begin{aligned}
E[R|Z=1] &= \int_\pi \big[\pi\big(p_a\tau(1,\mathbf{a}) + p_n(1-p_a)\overline{\theta}\tau(1,\mathbf{a}) + p_n(1-p_a)(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi},\mathbf{a}) + \\
&\quad (1-p_n)(1-p_a)\tau(\pi,\mathbf{a})\big) + \\
&\quad (1-\pi)\big(p_a\tau(0,\mathbf{a}) + (1-p_a)p_n\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi},\mathbf{a}) + (1-p_a)(1-p_n)\tau(\pi,\boldsymbol{a})\big)\big] d\pi \\
&= \frac{1}{2} \\
E[R|Z=0] &= \int_\pi \big[\pi\big(p_n\overline{\theta}\tau(1,\mathbf{a}) + p_n(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi},\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\big) + \\
&\quad (1-\pi)\big(p_n\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+1-\pi},\mathbf{a}) + (1-p_n)\tau(\pi,\boldsymbol{a})\big)\big] d\pi \\
&= \frac{1}{2}
\end{aligned}$$

- $q \in \Big[\max\{\frac{1}{\overline{\theta}}, \frac{2b(1-\pi\overline{\theta})}{\underline{\theta}(2b(1-\pi\overline{\theta})+p\overline{\theta}(1-\pi))}\}, \frac{1}{\underline{\theta}}\Big)$: In this region, the posterior belief upon realization of N is $\mu = \frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}}$ if public goods are observed, and $\mu = \frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})}$ if public goods are not observed. Realization of A provides no additional information since both types allocate to public goods, so the posterior is equivalent to the prior (if N is not realized) or the posterior (if N is realized). Note that in this case, $\tau(\cdot)$ accounts for the fact that the incompetent type will shirk in the second period.

$$\begin{aligned}
E[R|Z=1] = &\int_\pi \big[\pi\big(p_n\overline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + p_n(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + \\
&(1-p_n)\tau(\pi,\boldsymbol{a})\big) + (1-\pi)\big(p_n\underline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + \\
&p_n(1-\underline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\big)\big]d\pi \\
= &\int_\pi \frac{b-\underline{\theta}q(1-\pi)}{2b}d\pi = \frac{(2b-\underline{\theta}q(2-\pi))\pi}{4b} \\
E[R|Z=0] = &\int_\pi \big[\pi\big(p_n\overline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + p_n(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + \\
&(1-p_n)\tau(\pi,\boldsymbol{a})\big) + (1-\pi)\big(p_n\underline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + \\
&p_n(1-\underline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\big)\big]d\pi \\
= &\int_\pi \frac{b-\underline{\theta}q(1-\pi)}{2b}d\pi = \frac{(2b-\underline{\theta}q(2-\pi))\pi}{4b}
\end{aligned}$$

- $q \geq \frac{1}{\underline{\theta}}$. In this region, the posterior belief upon realization of N is $\mu = \frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}}$ if public goods are observed, and $\mu = \frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})}$ if public goods are not observed. Realization of A provides no additional information since both types allocate to public goods, so the posterior is equivalent to the prior (if N is not realized) or the posterior (if N is realized). This differs by from the previous case in the second period allocation of the incompetent type, which enters in $\tau(\cdot)$.

$$\begin{aligned}
E[R|Z=1] = & \int_\pi \big[\pi\big(p_n\overline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + p_n(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + \\
& (1-p_n)\tau(\pi,\boldsymbol{a})\big) + (1-\pi)\big(p_n\underline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + \\
& p_n(1-\underline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\big)\big]d\pi \\
= & \frac{1}{2} \\
E[R|Z=0] = & \int_\pi \big[\pi\big(p_n\overline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + p_n(1-\overline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + \\
& (1-p_n)\tau(\pi,\boldsymbol{a})\big) + (1-\pi)\big(p_n\underline{\theta}\tau(\frac{\pi\overline{\theta}}{\pi\overline{\theta}+(1-\pi)\underline{\theta}},\mathbf{a}) + \\
& p_n(1-\underline{\theta})\tau(\frac{\pi(1-\overline{\theta})}{\pi(1-\overline{\theta})+(1-\pi)(1-\underline{\theta})},\mathbf{a}) + (1-p_n)\tau(\pi,\mathbf{a})\big)\big]d\pi \\
= & \frac{1}{2}
\end{aligned}$$

First, I show that there is no evidence that incumbents in audited and un-audited municipalities run at different rates in Table C.13. Because the audits were rolled out in sequential waves (lotteries), I also estimate marginal effects of each lottery round (each municipality was audited only once) to ensure that the decision the ATE is not obscuring heterogeneity as a function of audit timing. All models include state fixed effects because each lottery effectively blocked on state and the sample comprises all municipalities with a population under 450,000. Following Hartman and Hidalgo (2018), a two one-sided t-test test rejects the null hypothesis of that $|ATE| \geq \pm 0.36\sigma$ at $p < 0.001$, where σ is the variance of the outcome.

To examine evidence that the ATE of audits on incumbent electoral performance, I employ equivalence testing as in Hartman and Hidalgo (2018). The theoretical prediction is that $ATE = 0$, so a traditional null hypothesis is an inappropriate test of the prediction. As such, I test the evidence against a null of an $|ATE| \geq 0.36\sigma$, a hypothesis recommended by Hartman and Hidalgo (2018). I examine both raw (i.e. difference-in-means) and covariate-adjusted specifications. In

	Incumbent contested re-election	
	(1)	(2)
Audited (indicator)	-0.03	
	(0.02)	
Audited, round 2		-0.02
		(0.11)
Audited, round 3		-0.07
		(0.07)
Audited, round 4		-0.02
		(0.07)
Audited, round 5		-0.06
		(0.07)
Audited, round 6		0.06
		(0.07)
Audited, round 7		-0.07
		(0.07)
Audited, round 8		-0.01
		(0.07)
Audited, round 9		-0.03
		(0.07)
Audited, round 10		-0.08
		(0.06)
Audited, round 11		0.06
		(0.07)
State FE	✓	✓
Mean, un-audited municipalities	0.426	0.426
Num. obs.	5496	5496

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.13: This table examines whether audited incumbents contest re-election at a differential rate than unaudited incumbents. Column 1 examines the ATE of audits on contesting re-election. Column 2 estimates the marginal effects of auditing in each round (a measure of treatment timing). The unit is the municipality. Heteroskedasticity-robust standard errors in parentheses.

principle, covariate-adjustment is needed to adjust for the blocking randomization strategy used in the CGU lotteries. I use two electoral dependent variables: an indicator for re-election and change in incumbent vote-share.

Figure C.13 rejects the null hypothesis of $|ATE| \geq 0.36\sigma$ at the $p < 0.005$ level in all specifications. We reject this null for all subsamples of bureaucratic quality except for the "re-elected" outcome in Tercile 2. Note however, that the power of the test is limited (analytically 0.63) for this specification. Moreover, there is no evidence of such an effect on change in vote share in the same tercile. This provides one way to test the prediction of an ATE of 0 in a frequentist framework.[3]

[3] An alternate test would use permutation tests/randomization inference. However, the assumption of a constant treatment effect for all units is inconsistent with this setting or the data where we expect (and Ferraz and Finan (2008) find) evidence of heterogeneity.

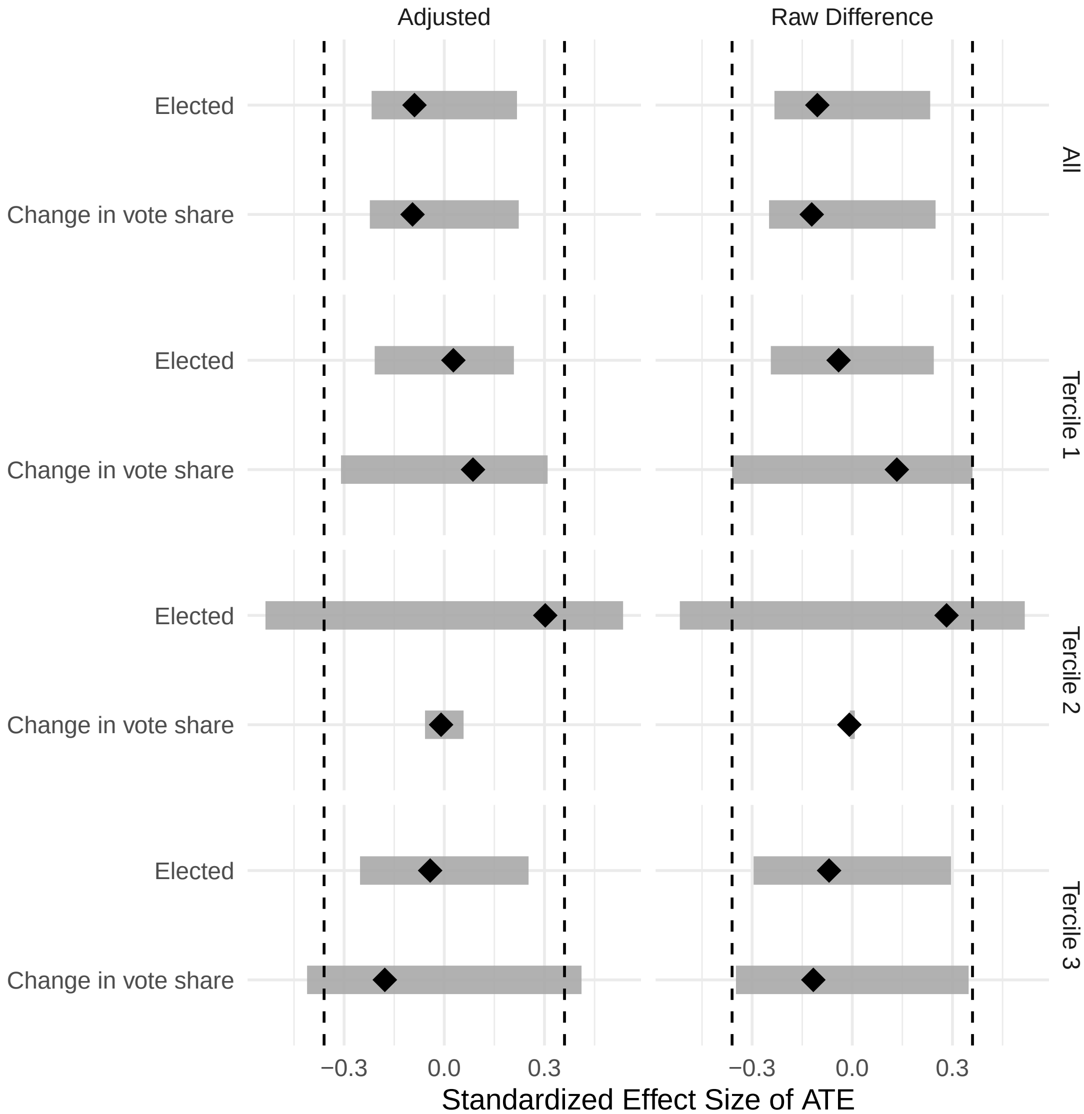


Figure C.13: Visualizations of two one-sided T-tests. The top panel presents the omnibus test across all municipalities where an incumbent contested re-election in 2004, $N = 2228$. The bottom three panels disaggregate by tercile. Points represent the standardized estimated ATE. The gray rectangles correspond to the inverted equivalence range that is ascertained from the data. The vertical lines represent the null hypothesis against which I am testing.

C.7 Existing studies of information and accountability

I identify 16 studies examining information and accountability for the purposes of Figure 4.4. Table C.14 provides the relevant citations.

	Country	Citation	Design	Metaketa-I	Included in Fig. 4.5
1	Benin	Adida et al. (2017)	E	✓	✓
2	Brazil	Ferraz and Finan (2008)	NE		
3	Brazil	Boas, Hidalgo, and Melo (2019)	E	✓	✓
4	Burkina Faso	Lierl and Holmlund (2019)	E	✓	✓
5	India	Banerjee et al. (2011)	E		✓
6	India	George, Gupta, and Neggers (2018)	E		✓
7	Philippines	Cruz, Keefer, and Labonne (2018)	E		
8	Philippines	Cruz et al. (2019)	E		✓
9	Mexico	Chong et al. (2015)	E		✓
10	Mexico	Arias et al. (2019)	E	✓	✓
11	Mexico	Enríquez et al. (2019)	E		✓
12	Mexico	Larreguy, Marshall, and Snyder Jr. (2020)	NE		
13	Senegal	Bhandari, Larreguy, and Marshall (2019)	E		✓
14	Uganda	Humphreys and Weinstein (2012)	E		
15	Uganda	Buntaine et al. (2018)	E	✓	✓
16	Uganda	Platas and Raffler (2019)	E	✓	✓

Table C.14: Studies of information and accountability and their locations. Under design, “E” corresponds to an experiment and “NE” corresponds to a natural experiment (one where the investigators did not manipulate provision of information).

Bibliography

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105 (8): 2364–2409.

Acemoglu, Daron, and James A. Robinson. 2008. "Persistence of Power, Elites, and Institutions." *American Economic Review* 98 (1): 267–293.

Acemoglu, Daron, Maria Angelica Bautista, Pablo Querubín, and James A. Robinson. 2008. "Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia". In *Institutions and Economic Performance*, ed. Elhanan Helpman. Cambridge: Harvard University Press pp. 181–245.

Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.

Adida, Claire L., Jessica Gottlieb, Eric Kramon, and Gwyneth H. McClendon. 2017. "Reducing or Reinforcing In-Group Preferences? An Experiment on Information and Ethnic Voting." *Quarterly Journal of Political Science* 12: 437–477.

Alesina, Alberto, and Guido Tabellini. 2007. "Bureaucrats or Politicians? Part I: A Single Policy Task." *American Economic Review* 97 (1): 169–179.

Arias, Eric, Horacio Larreguy, John Marshall, and Pablo Querubin. 2019. "Priors Rule: When do Malfeasance Revelations Help or Hurt Incumbent Parties?" Available at `https://scholar.harvard.edu/files/jmarshall/files/mexico_accountability_experiment_v13.pdf`.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review* 110 (5): 1355–1394.

Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15: 183–201.

Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2015. "All Else Equal in Theory and Data (Big or Small)." *PS Political Science* 48 (1): 89–94.

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Fridenberg. 2017. "Accountability and Information in Elections." *American Economic Journal: Microeconomics* 9 (2): 95–128.

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Fridenberg. 2018. "Learning about Voter Rationality." *American Journal of Political Science* 62 (1): 37–54.

Avis, Eric, Claudio Ferraz, and Frederico Finan. 2018. "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians." *Journal of Political Economy* 126 (5): 1912–1964.

Ba, Bocar. 2018. "Going the Extra Mile: the Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago." Working paper, University of California Irvine.

Banerjee, Abhijit. 1997. "A Theory of Misgovernance." *Quarterly Journal of Economics* 112 (3): 1289–1332.

Banerjee, Abhijit, Selvan Kumar, Rohini Pande, and Felix Su. 2011. "Do Informed Voters Make Better Choices? Experimental Evidence From India." Available at `https://scholar.harvard.edu/files/rpande/files/do_informed_voters_make_better_choices.pdf`.

Banks, Jeffrey S. 1989. "Agency Budgets, Cost Information, and Auditing." *American Journal of Political Science* 33 (3): 670–699.

Bardhan, Pranab. 2002. "Decentralization of Governance and Development." *Journal of Economic Perspectives* 16 (4): 185–205.

Becker, Gary S. 1957. *The Economics of Discrimination*. 2nd edition ed. Chicago, IL: University of Chicago Press.

Berwick, Elissa, and Fotini Christia. 2018. "State Capacity Redux: Integrating Classical and Experimental Contributions to an Enduring Debate." *Annual Review of Political Science* 21: 71–91.

Besley, Timothy, and Marta Reynal-Querol. 2011. "Do Democracies Select More Educated Leaders?" *American Political Science Review* 105 (3): 552–566.

Besley, Timothy, and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78 (1): 1–34.

Bhandari, Abhit, Horacio Larreguy, and John Marshall. 2019. "Able and Mostly Willing: An Empirical Anatomy of Information's Effect on Voter-Driven Accountability in Senegal." Available at `https://scholar.harvard.edu/files/jmarshall/files/accountability_senegal_paper_v5.pdf`.

Bhavnani, Rikhil R., and Alexaner Lee. 2018. "Local Embeddedness and Bureaucratic Performance: Evidence from India." *Journal of Politics* 80 (1): 71–87.

Boas, Taylor C., F. Daniel Hidalgo, and Marcus André Melo. 2019. "Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil." *American Journal of Political Science* 63 (2): 385–400.

Brierly, Sarah. 2020. "Unprincipled Principals: Co-opted Bureaucrats and Corruption in Ghana." *American Journal of Political Science* 64 (2).

Bueno de Mesquita, Ethan, and Scott A. Tyson. 2019. "The Commensurability Problem: Conceptual Difficulties in Estimating the Effect of Behavior on Behavior." Working paper.

Buntaine, Mark T., Ryan Jablonski, Daniel L. Nielson, and Paula M. Pickering. 2018. "SMS Texts on Corruption Help Ugandan Voters Hold Elected Councillors Accountable at the Polls." *Proceedings of the National Academy of Sciences* 115 (26): 6668–6673.

Bussell, Jennifer. 2019. *Clients and Constituents: Political Responsiveness in Patronage Democracies*. New York: Oxford University Press.

Calvo, Ernesto, and Maria Victoria Murillo. 2004. "Who Delivers? Partisan Clients in the Argentine Electoral Market." *American Journal of Political Science* 48 (4): 742–757.

Camacho, Adriana, and Emily Conover. 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy* 3: 41–65.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York, NY: John Wiley and Sons.

Centeno, Miguel Angel. 2003. *Blood and Debt: War and the Nation-State in Latin America*. University Park, PA: Penn State University Press.

Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2015. "Sources of Authoritarian Responsiveness: A Field Experiment in China." *American Journal of Political Science* 60 (2): 383–400.

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *The Quarterly Journal of Economics* 1902 (2): 179–221.

Chong, Alberto, Ana de la O, Dean Karlan, and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification." *Journal of Politics* 77 (1): 51–77.

Christensen, Darin, and Simon Ejdemyr. 2020. "Do Elections Improve Constituency Responsiveness? Evidence from US Cities." *Political Science Research and Methods* Forthcoming.

Clark, William Roberts, and Matt Golder. 2015. "Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?" *PS Political Science* 48 (1): 65–70.

Coppock, Alexander. 2018. "Avoiding Post-Treatment Bias in Audit Experiments." *Journal of Experimental Political Science* . Forthcoming.

Cruz, Cesi, Philip Keefer, and Julien Labonne. 2018. "Buying Informed Voters: New Effects of Information on Voters and Candidates." Available at `https://static1.squarespace.com/static/58c979fad1758e09d030809c/t/5c048e82898583120b1f73cc/1543802523246/buying_informed_voters_web.pdf`.

Cruz, Cesi, Philip Keefer, Julien Labonne, and Francesco Trebbi. 2019. "Making Policies Matter: Voter Responses to Campaign Promises." Working paper available at `https://static1.squarespace.com/static/58c979fad1758e09d030809c/t/5cfed616d6104500019dff1b/1560204824899/making_promises_matter_6102019.pdf`.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics* 128 (3): 1169–1218.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Who Becomes a Politician?" *Quarterly Journal of Economics* 132 (4): 1877–1914.

Dargent, Eduardo. 2016. *Tchnocracy and Democracy in Latin America*. New York: Cambridge University Press.

De la O, Ana. 2015. *Crafting Policies to End Poverty in Latin America*. New York: Cambridge University Press.

Defensoría del Pueblo de Colombia. 2019. La tutela y los derechos a la salud y a la seguridad social. Policy report Bogotá Defensoría del Pueblo de Colombia.

Dimitrov, Martin K. 2013. "What the Party Wanted to Know: Citizen Complaints as a "Barometer of Public Opinion" in Communist Bulgaria." *East European Politics and Societies: and Cultures* 28 (2): 271–295.

Dipoppa, Gemma, and Guy Grossman. 2020. "The Effect of Election Proximity on Government Responsiveness and Citizens' Participation: Evidence from English Local Elections." *Comparative Political Studies* Forthcoming.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretive Review." *Journal of Human Resources* 37 (4): 696–727.

Dixit, Avinash, and John Londregan. 1995. "Redistributive Polticis and Economic Efficiency." *American Political Science Review* 89 (4): 856–866.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis, eds. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. New York: Cambridge University Press.

Eframidis, Pavlos S., and Paul G. Spirakis. 2006. "Weighted Random Sampling with a Reservoir." *Information Processing Letters* 97 (5): 181–185.

Einstein, Katherine Levine, and David Glick. 2016. "Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats." *American Journal of Political Science* 61 (1): 100–116.

Enriquez, Elaine, and Miguel Angel Centeno. 2012. "State Capacity: Utilization, Durability, and the Role of Wealth vs. History." *RIMCIS-International and Multidisciplinary Journal of Social Sciences* 1 (2): 130–162.

Enríquez, José Ramón, Horacio Larreguy, John Marshall, and Alberto Simpser. 2019. "Information saturation and electoral accountability: Experimental evidence from Facebook in Mexico." Working paper.

Epstein, David, and Sharyn O'Halloran. 1994. "Administrative Procedures, Information, and Agency Discretion." *American Journal of Political Science* 38 (3): 697–722.

Evans, Peter B., and James E. Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth." *American Sociological Review* 64 (5): 748–765.

Fearon, James. 1999. *Democracy, Accountability, and Representation*. Cambridge University Press chapter Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance.

Fergusson, Leopoldo, Pablo Querubín, Nelson A. Ruiz, and Juan F. Vargas. 2018. "The Real Winner's Curse." Working paper.

Ferrali, Romain, and Galileu Kim. 2019. "Are Anti-Corruption Audits Effective? Evidence From Brazilian Municipal Bureaucracies." Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3499117.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123 (2): 703–745.

Ferraz, Claudio, and Frederico Finan. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101 (4): 1274–1311.

Fiszbein, Ariel, Norbert Schady, Francisco H.G. Ferrerira, Margaret Grosh, Niall Keleher, Pedro Olinto, and Emmanuel Skoufias. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington DC: The International Bank for Reconstruction and Development/The World Bank.

Fizbein, Ariel. 1997. "The Emergence of Local Capacity: Lessons From Colombia." *World Development* 25 (7): 1029–1043.

Folke, Olle, Shigeo Hirano, and James M Snyder. 2011. "Patronage and Elections in U.S. States." *American Political Science Review* 105 (3): 567–585.

Fowler, Anthony, and Andrew B. Hall. 2018. "Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence." *Journal of Politics* 80 (4): 1423–1437.

Fowler, Anthony, and B. Pablo Montagnes. 2015. "College Football, Elections, and False-Positive

Results in Observational Research." *Proceedings of the National Academy of Sciences (PNAS)* 112 (45): 13800–13804.

Fox, Justin, and Stuart V. Jordan. 2011. "Delegation and Accountability." *Journal of Politics* 73 (3): 831–844.

Gaikwad, Nikhar, and Gareth Nellis. 2017. "Do Politicians Discriminate Against Internal Migrants? Evidence from Nationwide Field Experiments in India." Working paper.

Gailmard, Sean, and John W. Patty. 2007. "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise." *American Journal of Political Science* 51 (4): 873–889.

Gailmard, Sean, and John W. Patty. 2012. "Formal Models of Bureaucracy." *Annual Review of Political Science* 15: 353–377.

Garay, Candelaria. 2016. *Social Policy Expansion in Latin America*. New York: Cambridge University Press.

Geddes, Barbara. 1991. "A Game Theoretic Model of Reform in Latin American Democracies." *American Political Science Review* 85 (2): 371–392.

Geddes, Barbara. 1994. *Politician's Dilemma: Building State Capacity in Latin America*. Berkeley, CA: University of California Press.

Gennaioli, Nicola, and Hans-Joachim Voth. 2015. "State Capacity and Military Conflict." *Review of Economic Studies* 82 (1409-1448).

George, Siddharth, Sarika Gupta, and Yusuf Neggers. 2018. "Coordinating Voters against Criminal Politicians: Evidence from a Mobile Experiment in India." Available at `https://scholar.harvard.edu/files/siddharthgeorge/files/voter_mobile_experiment_181126.pdf`.

Golden, Miriam, and Brian Min. 2013. "Distributive Politics Around the World." *Annual Review of Political Science* 16: 73–99.

Gonçalves, Sónia. 2013. "The Effects of Participatory Budgeting on Municipal Expendituresand Infant Mortality in Brazil." *World Development* 53: 94–110.

González González, Fernán E. 2014. *Poder y Violencia en Colombia*. Bogotá, D.C.: ODECOFI-CINEP.

Grindle, Merilee S. 2012. *Jobs for the Boys: Patronage and the State in Comparative Perspective*. Cambridge, MA: Harvard University Press.

Hamel, Brian T., and Derek E. Holliday. 2019. "Who Gets a Streetlight? Unequal Responsiveness in City Services." Working paper, University of California Los Angeles.

Hartman, Erin, and F. Daniel Hidalgo. 2018. "An Equivalence Approach to Balance and Placebo Tests." *American Journal of Political Science* 62 (4): 1000–1013.

Healy, Andrew J., Neil Malhotra, and Cecilia Hyunjung Mo. 2010. "Irrelevant Events Affect Voters' Evaluations of Government Performance." *Proceedings of the National Academy of Sciences (PNAS)* 107 (29): 12804–12809.

Hemker, Johannes, and Anselm Rink. 2017. "Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices." *American Journal of Political Science* Forthcoming.

Hicken, Allen, and Noah L. Nathan. 2020. "Clientelism's Red Herrings: Dead Ends and New

Directions in the Study of Nonprogrammatic Politics." *Annual Review of Political Science* 23: 277–294.

Hirschman, Alberto O. 1970. "The Search for Paradigms as a Hindrance to Understanding." *World Politics* 22 (3): 329–343.

Holland, Alisha C. 2015. "The Distributive Politics of Enforcement." *American Journal of Political Science* 59 (2): 357–371.

Holland, Alisha C. 2017. *Forbearance as Redistribution: The Politics of Informal Welfare in Latin America*. New York, NY: Cambridge University Press.

Huber, Evelyne, and John D. Stephens. 2012. *Democracy and the Left: Social Policy and Inequality in Latin America*. Chicago, IL: University of Chicago Press.

Huber, John D. 2017. *Exclusion by Elections: Inequality, Ethnic Identity, and Democracy*. New York: Cambridge University Press.

Huber, John D., and Charles R. Shipan. 2002. *Deliberate Discretion? The Institutional Foundations of Bureaucratic Autonomy*. New York: Cambridge University Press.

Huber, John D., and Michael M. Ting. 2020. "Civil Service and Patronage in Bureaucracies." *Journal of Politics* Forthcoming.

Huber, John D., and Nolan McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98 (3): 481–494.

Humphreys, Macartan, and Jeremy M. Weinstein. 2012. "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda - Preliminary Analysis." IGC Working Paper S-5021-UGA-1.

Ibáñez, Ana María, and Carlos Eduardo Vélez. 2008. "Civil Conflict and Forced Migration: The Micro Determinants and Welfare Losses of Displacement in Colombia." *World Development* 36 (4): 659–676.

Incerti, Trevor. 2019. "Corruption Information and Vote Share: A Meta-Analysis and Lessons for Experimental Design." Working paper, available at `https://tincerti.github.io/files/corruption_voting_meta.pdf`.

Izzo, Federica, Torun Dewan, and Stephane Wolton. 2019. "Cumulative Knowledge in the Social Sciences: The Case of Improving Voters' Information." Working paper.

Jia, Ruixue, Masayuki Kudamatsu, and David Seim. 2015. "Political Selection in China: The Complementary Roles of Connections and Performance." *Journal of the European Economic Association* 13 (4): 631–668.

Johnson, Noel D., and Mark Koyama. 2017. "States and Economic Growth: Capacity and Constraints." *Explorations in Economic History* 64: 1–20.

Kasara, Kimuli, and Pavithra Suryanarayan. 2015. "When Do the Rich Vote Less Than the Poor and Why? Explaining Turnout Inequality across the World." *American Journal of Political Science* 59 (3): 613–627.

Klašnja, Marko, and Rocío Titunik. 2017. "The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability." *American Political Science Review* 111 (1): 129–148.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. "The Bias is Built In: How Administrative Records Mask Racially Biased Policing." *American Political Science Review* Forthcoming.

Kruks-Wisner, Gabrielle. 2018. *Claiming the State: Active Citizenship and Social Welfare*. New York: Cambridge University Press.

Larreguy, Horacio, John Marshall, and James M. Snyder Jr. 2020. “Publicizing Malfeasance: When the Local Media Structure Facilitates Electoral Accountability in Mexico.” *Economic Journal* Forthcoming.

Lee, Melissa M., and Nan Zhang. 2016. “Legibility and the Informational Foundations of State Capacity.” *Journal of Politics* Journal of Politics (79): 1.

Leff, Nathaniel H. 1964. “Economic Development Through Bureaucratic Corruption.” *American Behavioral Scientist* 8 (3): 8–14.

Lewis, David. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton, NJ: Princeton University Press.

Li, Christopher M., Greg Sasso, and Ian Turner. 2019. “Accountability in Governing Hierarchies.” Working paper.

Lichand, Guilherme, Marcos F M Lopes, and Marcelo Medeiros. 2016. “Is Corruption Good for Your Helth?” Working paper.

Lierl, Malte, and Marcus Holmlund. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Cambridge University Press chapter Performance Information and Voting Behavior in Burkina Faso’s Municipal Elections: Separating the Effects of Information Content and Information Delivery, pp. 221–256.

Lin, Winston. 2013. “Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique.” *Annals of Applied Statistics* 7 (1): 295–318.

Lindbeck, Assar, and Jorgen W. Weibull. 1987. “Balanced-Budget Redistribution as the Outcome of Political Competition.” *Public Choice* 52 (3): 273–297.

Lipsky, Michael. 2010. *Street-Level Democracy: Dilemmas of the Individual in Public Service*. 2nd edition ed. New York, NY: Russell Sage Foundation.

Magaloni, Beatriz, Edgar Franco-Vivanco, and Vanessa Mello. 2020. "Killing in the Slums: Social Order, Criminal Governance, and Police Violence in Rio de Janeiro." *American Political Science Review* 114 (2): 552–572.

Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms, and Results." *European Journal of Sociology* 25 (2): 185–213.

Mares, Isabela, and Lauren Young. 2018. "The Core Voter's Curse: Clientelistic Threats and Promises in Hungarian Elections." *Comparative Political Studies* 51 (11): 1441–1471.

Marshall, John. 2019. "When Can Close Election RDDs Identify the Effects of Winning Politician Characteristics?" Working paper, Columbia University.

Martin, Lucy, and Pia J. Raffler. 2019. "Fault Lines: The Effects of Bureaucratic Power on Electoral Accountability." *American Journal of Political Science* Forthcoming.

Martine, George. 1975. "Volume, Characteristics and Consequences of Internal Migration in Colombia." *Demography* 12 (2): 193–208.

Martz, John D. 1997. *The Politics of Clientelism in Colombia*. New Brunswick, NJ: Transaction Publishers.

Maskin, Eric, and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government." *American Economic Review* 94 (4): 1034–1054.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110 (3): 681–712.

Mayka, Lindsay R. 2016. "Colombia's Surprising Resilience." *Journal of Democracy* 27 (3): 139–147.

McCarty, Nolan. 2004. "The Appointments Dilemma." *American Journal of Political Science* 48 (3): 413–428.

McCubbins, Matthew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28 (1): 165–179.

Misión de Observación Electoral. 2018. Mapas y factores de resgo electoral: Elecciones nacionales Colombia 2018. Technical report Misión de Observación Electoral Bogotá, D.C.: .

Niedzwiecki, Sara. 2018. *Uneven Social Policies: The Politics of Subnational Variation in Latin America*. New York: Cambridge University Press.

Ocampo, Gloria Isabel. 2014. *Poders Regionales, Clientelismo, y Estado: Etnografía del Poder y La Política en Córdoba, Colombia*. Bogotá, D.C.: ODECOFI-CINEP.

Pan, Jennifer, and Kaiping Chen. 2018. "Concealing Corruption: How Chinese Officials Distort Upward Reporting of Online Grievances." *American Political Science Review* 112 (3): 602–620.

Parker, Susan W., and Petra E. Todd. 2017. "Conditional Cash Transfers: The Case of Progresa/Oportunidades." *Journal of Economic Literature* 55 (3): 866–915.

Pepinsky, Thomas B., Jan H. Pierskalla, and Audrey Sacks. 2017. "Bureaucracy and Service Delivery." *Annual Review of Political Science* 20: 249–268.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–661.

Platas, Melina R., and Pia Raffler. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Number 6 New York: Cambridge University Press chapter Candidate Videos and Vote Choice in Ugandan Parliamentary Elections, pp. 156–187.

Portes, Alejandro, and Kelly Hoffman. 2003. "Latin American Class Structures: Their Composition and Change during the Neoliberal Era." *Latin American Research Review* 38 (1): 41–82.

Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111 (5): 929–958.

Prendergast, Canice. 2007. "The Motivation and Bias of Bureaucrats." *American Economic Review* 96 (1): 180–196.

Przeworski, Adam, Susan Stokes, and Bernard Manin. 1999. *Democracy, Accountability, and Representation*. Cambridge University Press chapter Introductoin, pp. 1–26.

Querubin, Pablo. 2016. "Family and Politics: Dynastic Persistence in the Philippines." *Quarterly Journal of Political Science* 11 (2): 151–181.

Raffler, Pia. 2019. "Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy." Working paper, Harvard University.

Rasul, Imran, and Daniel Rogger. 2016. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* forthcoming.

Rauch, James E. 1995. "Bureaucracy, Infrastructure, and Economic Growth: Evidence from U.S. Cities During the Progressive Era." *American Economic Review* 85 (4): 968–979.

Rivera, Roman G., and Bocar A. Ba. 2019. "The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago." University of California Los

Angeles, available at https://www.dropbox.com/s/4ar67esosdcuiuu/Rivera%20and%20Ba%20%282019%29.pdf?dl=0.

Rizzo, Tesalia. 2018. "When Clients Exit: Breaking the Clientelistic Feedback Loop." Working paper.

Rizzo, Tessalia. 2019. "When Clients Exit: Breaking the Clientelist Feedback Loop." Working paper.

Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *Journal of Politics* 78 (3): 941–955.

Sanders, James E. 2004. *Contentious Republicans: Popular Politics, Race, and Class in Nineteenth-Century Colombia*. Durham, NC: Duke University Press.

Schmidt, Steffen W. 1974. "Bureaucrats as Modernizing Brokers? Clientelism in Colombia." *Comparative Politics* 6 (3): 425–450.

Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.

Seabra, Sérgio N. 2018. "How Auditing Can be Directed Against Corruption: The Case of Brazil's Municipality-Facing Auditing Program." Working paper, University of Oxford.

Sjoberg, Fredrik M., Jonathan Mellon, and Tiago Peixoto. 2017. "The Effect of Bureaucratic Responsiveness on Citizen Participation." *Public Administration Review* 77 (3): 340–351.

Slough, Tara. 2019. "On Theory and Identification: When and Why We Need Theory for Causal Identification." Working paper, Columbia University. Available at www.taraslough.com/assets/pdf/theory_id.pdf.

Slough, Tara, and Christopher Fariss. 2020. “Misgovernance and Human Rights: The Case of Illegal Detention without Intent.” *American Journal of Political Science* Forthcoming.

Smith, Daniel M. 2018. *Dynasties and Democracy: The Inherited Incumbency Advantage in Japan*. Stanford, CA: Stanford University Press.

Snowberg, Erik, and Michael M. Ting. 2019. “An Organizational Theory of State Capacity.” Working paper.

Soares, Fábio Veras, Raphael Perez Ribas, and Rafael Gurreiro Osório. 2010. “Evaluating the Impact of Brazil’s Bolsa Família: Cash Transfer Programs in Comparative Perspective.” *Latin American Research Review* 45 (2): 173–190.

Soifer, Hillel David. 2015. *State Building in Latin America*. New York: Cambridge University Press.

Streeter, Shea. 2019. “Lethal Force in Black and White: Assessing Racial Disparities in the Circumstances of Police Killings.” *Journal of Politics* 81 (3): 1124–1132.

Suryanarayan, Pavithra. 2020. “Hollowing out the State: Franchise Expansion and Fiscal Capacity in Colonial India.” Working paper, available at `https://pavisuridotcom.files.wordpress.com/2012/08/pavisuri_capacity.pdf`.

Taylor, Whitney. 2018. “Ambivalent Legal Mobilization: Perceptions ofJustice and the Use of the Tutela in Colombia.” *Law and Society Review* 52 (2).

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990-1992*. Malden, MA: Blackwell.

Ting, Michael M. 2017. “Politics and Administration.” *American Journal of Political Science* 61 (2): 305–319.

Ting, Michael M, James M Snyder Jr., Shigeo Hirano, and Olle Folke. 2012. “Elections and Reform: The Adoption of Civil Service Systems in the U.S. States.” *Journal of Theoretical Politics* 25 (3): 363–387.

Toral, Guillermo. 2019. “The benefits of patronage: How the political appointment of bureaucrats can enhance their accountability and effectiveness.” Working paper, MIT.

Ujhelyi, Gergely. 2014. “Civil Service Rules and Policy Choices: Evidence from US State Governments.” *American Economic Journal: Economic Policy* 6 (2): 338–380.

Uribe de Hincapié, María Tereesa, and Jesús María Álvarez Gaviria. 1998. *Las Raíces del Poder Regional: El Caso Antioqueño*. Medellín: Editorial Universidad de Antioquia.

Van Rijckeghem, Caroline, and Beatrice Weder. 2001. “Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much?” *Journal of Development Economics* 65 (2): 307–331.

Weber, Eugun. 1976. *Peasants Into Frenchmen: The Modernization of Rural France*. Palo Alto, CA: Stanford University Press.

Weitz-Shapiro, Rebecca. 2012. “What Wins Votes: Why Some Politicians Opt Out of Clientelism.” *American Journal of Political Science* 56 (3): 568–583.

Weitz-Shapiro, Rebecca, and Matthew S. Winters. 2016*a*. “Can Citizens Discern? Information Credibility, Political Sophistication, and the Punishment of Corruption in Brazil.” *Journal of Politics* 79 (1): 60–74.

Weitz-Shapiro, Rebecca, and Matthew S. Winters. 2016*b*. "“Can Citizens Discern? Information Credibility, Political Sophistication, and the Punishment of Corruption in Brazil” Online Appendix." Supplementary information available at `https://www.journals.uchicago.edu/doi/suppl/10.1086/687287/suppl_file/150469Appendix.pdf`.

White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109 (1): 129–142.

Williams, Martin J. 2017. "The Political Economy of Unfinished Development Projects: Corruption, Clientelism, or Collective Choice?" *American Political Science Review* 111 (4): 705–723.

Winters, Matthew S., and Rebecca Weitz-Shapiro. 2016. "Who's in Charge Here? Direct and Indirect Accusations and Voter Punishment of Corruption." *Political Research Quarterly* 69 (2): 207–219.

Wood, Elisabeth Jean, Douglas Rogers, K. Sivaramakrishnan, and Rene Almeling. 2020. "Resuming Field Research in Pandemic Times." Social Science Research Council, available at `https://items.ssrc.org/covid-19-and-the-social-sciences/social-research-and-insecurity/resuming-field-research-in-pandemic-times/`.

Yazaki, Yukihiro. 2018. "The Effects of Bureaucracy on Political Accountability and Electoral Selection." *European Journal of Political Economy* 51: 57–68.